

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:12:44 ; Search time 200 Seconds  
(without alignments)  
2926.263 Million cell updates/sec

Title: US-10-736-769-4  
Perfect score: 6909  
Sequence: 1 MAAGAGRGWLLWALLRLAQ.....GSIKGAGATSNFLPNNGRQF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_21.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	8	ADJ27297 Human NPC
2	6909	100.0	1332	9	ADY60740 Human Nie
3	6909	100.0	1332	9	AEB93570 Human NPC
4	6896	99.8	1332	4	AAM79169 Human pro
5	6896	99.8	1332	4	AAG5638 SSD-conta
6	6872.5	99.5	1359	4	AAM79168 Human pro
7	6872.5	99.5	1359	8	ADJ27337 Human NPC
8	6872.5	99.5	1359	9	ADY60780 Human Nie
9	6872.5	99.5	1359	9	AEB93610 Human NPC
10	6536	94.6	1344	4	ABG22693 Novel hum
11	5421.5	78.5	1331	8	ADJ27295 Rat NPC1L
12	5421.5	78.5	1331	9	ADY60738 Rat Niema
13	5421.5	78.5	1331	9	AEB93568 Rat NPC1L
14	5407	78.3	1333	8	ADJ27305 Mouse NPC
15	5407	78.3	1333	9	ADY60748 Mouse Nie
16	5407	78.3	1333	9	AEB93578 Mouse NPC
17	4466	64.6	982	4	ABG22691 Novel hum
18	2402.5	34.8	1278	2	AAW88445 Human NPC
19	2402.5	34.8	1278	3	AAB42983 Human ORF
20	2402.5	34.8	1278	8	ADQ39879 Human myo
21	2402.5	34.8	1278	8	ADU06723 Novel bro
22	2385	34.5	1319	2	AAW88446 Mouse NPC
23	2262	32.7	1287	4	ABB61737 Drosophil
24	2262	32.7	1287	8	ADS96670 Drosophil

25	1864.5	27.0	1223	4	ABBS58629	Abb58629 Drosophil
26	1329	19.2	1170	2	AAW88447	Aaw88447 Yeast NPC
27	1078	15.6	1296	2	AAW88448	Aaw88448 Caenorhab
28	963	13.9	229	4	AAM80153	Aam80153 Human pro
29	963	13.9	229	4	AAM80152	Aam80153 Human pro
30	942	13.6	194	4	AAG5637	Aag5637 SSD domai
31	651.5	9.4	1447	2	AAW75375	Aar75375 Human pat
32	651.5	9.4	1447	2	AAW52200	Aaw52200 Human pat
33	651.5	9.4	1447	2	AAW72969	Aaw72969 Human pat
34	651.5	9.4	1447	4	AAB67163	Aab67163 Human pat
35	651.5	9.4	1447	5	AAE19830	Aae19830 Human pat
36	651.5	9.4	1447	5	ABJ10931	Abj10931 TRC8 rela
37	651.5	9.4	1447	5	AAG79571	Aag79571 Human pat
38	651.5	9.4	1447	7	ABU62275	Abu62275 Human pat
39	651.5	9.4	1447	7	ADD46678	Ad46678 Human Pro
40	651.5	9.4	1447	7	ADG94224	Adg94224 Human ptc
41	651.5	9.4	1447	7	ADH62731	Adh62731 Human pat
42	651.5	9.4	1447	8	ADG48989	Ade48989 Human pat
43	646.5	9.4	1434	2	AAW52199	Aaw52199 Mouse pat
44	646.5	9.4	1434	2	AAW72968	Aaw72968 Mouse pat
45	646.5	9.4	1434	4	AAB67159	Aab67159 Murine pa

ALIGNMENTS

RESULT 1

ADJ27297	
ID	ADJ27297 standard; protein; 1332 AA.
XX	
AC	ADJ27297;
XX	
DT	20-MAY-2004 (first entry)
XX	
DE	Human NPC1L1.
XX	
KW	Niemann-Pick disease; type Cl; gene-like 1; NPC1L1; trans-golgi network;
KW	plasma membrane; transport signal; promoter;
KW	sterol regulated element binding protein 1; SREBP1;
KW	binding consensus sequence; transmembrane domain; sterol-sensing domain;
KW	SSD; cholesterol; NPC1; receptor; Niemann-Pick Cl disease; intestinal;
KW	cholesterol absorption; serum cholesterol; hyperlipidaemia;
KW	atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
XX	
OS	Homo sapiens.
XX	
PN	WO2004009772-A2.
XX	
PD	29-JAN-2004.
XX	
PF	17-JUL-2003; 2003WO-US022467.
XX	
PR	19-JUL-2002; 2002US-0397442P.
XX	
PA	(SCHE ) SCHERING CORP.
XX	
PI	Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;
XX	
DR	WPI; 2004-132945/13.
DR	N-PSDB; ADJ27296.
DR	GENBANK; AF192522.
XX	
PT	New Niemann-Pick disease, type Cl, gene-like 1 polypeptide, for detecting
PT	or identifying antagonists of NPC1L1 for inhibiting intestinal
PT	cholesterol absorption in a subject, or for treating elevated serum
PT	cholesterol or stroke.
XX	
PS	Claim 11; SEQ ID NO 4; 125pp; English.
XX	
CC	This sequence represents a Niemann-Pick disease, type Cl, gene-like 1
CC	(NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains
CC	a motif ADJ27331 which acts as a trans-golgi network to plasma membrane
CC	transport signal, and which exhibits limited tissue distribution and



XX Novel isolated NPC1l1 polypeptide, useful for identifying antagonist of  
PT NPC1l1 that inhibits NPC1l1 mediated sterol or 5 alpha-sterol uptake, for  
PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.  
XX Claim 30; SEQ ID NO 4; 146pp; English.  
XX  
CC The invention relates to an isolated Niemann-Pick C1-like protein 1  
CC (NPC1l1) polypeptide comprising 42 or more contiguous amino acids from an  
CC amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,  
CC human and Mouse NPC1l1). Also included are an isolated polynucleotide  
CC encoding NPC1l1 (appearing as ADY60737, ADY60739 or ADY60747), a  
CC recombinant vector comprising the polynucleotide, a host cell comprising  
CC the vector, an isolated antibody which specifically binds to NPC1l1 (or  
CC to a rat NPC1l1 peptide sequence appearing as ADY60775-ADY60778),  
CC preparation of NPC1l1, a mutant transgenic mouse comprising a homozygous  
CC mutation of endogenous chromosomal NPC1l1 (where the mouse does not  
CC produce any functional NPC1l1 protein), an offspring or progeny of the  
CC mouse having inherited a mutated NPC1l1 allele, an offspring or progeny of the  
CC sample for an intestinal sterol or Salpha-sterol absorption, screening a  
CC inhibiting NPC1l1 mediated sterol or Salpha-sterol absorption antagonist,  
CC kit (comprising ezetimibe or substituted azetidinone in a pharmaceutical  
CC dosage form, and information indicating that NPC1l1 is a target of  
CC ezetimibe or substituted azetidinone), decreasing the level of intestinal  
CC sterol or Salpha-sterol absorption in a subject (involving reducing the  
CC level of expression of NPC1l1 in the subject), identifying an antagonist  
CC of NPC1l1 and an isolated mammalian cell which lacks a gene encoding a  
CC functional NPC1l1 protein. NPC1l1 is useful for identifying an antagonist  
CC of NPC1l1 which is useful for inhibiting or decreasing the level of  
CC NPC1l1 mediated sterol or Salpha-sterol uptake in a subject such as  
CC mouse, rat or human. The level of expression of NPC1l1 in the subject is  
CC reduced by mutating NPC1l1 in the subject. The NPC1l1 knockout mouse is  
CC useful for screening a sample for intestinal sterol or Salpha-sterol  
CC absorption antagonist. The NPC1l1 antagonists identified are useful for  
CC treating medical conditions caused or mediated by NPC1l1, e.g.,  
CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or  
CC arteriosclerosis. The present sequence represents a mammalian NPC1l1  
XX protein.  
XX  
SQ Sequence 1332 AA;

Query Match 100.0%; Score 6909; DB 9; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEAGLGRWLLWALLRLAQSEPTTHQPGYCAFYDECGKPELSGSLMTLSNVSCLSN 60  
DB 1 MAEAGLGRWLLWALLRLAQSEPTTHQPGYCAFYDECGKPELSGSLMTLSNVSCLSN 60  
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVLSLEASLITKALLTRCPACSDNF 120  
DB 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVLSLEASLITKALLTRCPACSDNF 120  
QY 121 VNLHCHTCSNQSLEFINTRVNQLGAGQLPAVVAEAFYQHSFABQSDSCSRVRVPA 180  
DB 121 VNLHCHTCSNQSLEFINTRVNQLGAGQLPAVVAEAFYQHSFABQSDSCSRVRVPA 180  
QY 181 ATLAVCTMGVYGSALCNQRMWLFQDGTGNGLAPLDITPHLLERQAVGSGIQLNEGV 240  
DB 181 ATLAVCTMGVYGSALCNQRMWLFQDGTGNGLAPLDITPHLLERQAVGSGIQLNEGV 240  
QY 241 ARCNEQSGDDVATCSQDCAACPAIARPAQALDSTFYLQMPGSLVLIILICSVFAVVTI 300  
DB 241 ARCNEQSGDDVATCSQDCAACPAIARPAQALDSTFYLQMPGSLVLIILICSVFAVVTI 300  
QY 301 LLVGFVRVARDKSKMVDPKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVLSV 360  
DB 301 LLVGFVRVARDKSKMVDPKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVLSV 360  
QY 361 IPWVLAAGLVFTLTTDPELWSAPNSQARSEKAFHDQHFQFFRTNQVILTAPNRSSY 420  
DB 361 IPWVLAAGLVFTLTTDPELWSAPNSQARSEKAFHDQHFQFFRTNQVILTAPNRSSY 420

QY 421 RYDSSLGPKNFGSILDLLELLELQERLRLHQLVMSPEAQRNLSIQDICYAPLNPNT 480  
DB 421 RYDSSLGPKNFGSILDLLELLELQERLRLHQLVMSPEAQRNLSIQDICYAPLNPNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFDKGTALAL 540  
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFDKGTALAL 540  
QY 541 SCWADYGAPVFPPLAIGGYKGDYSBAEALIMTFSLNNYPAGPRLAQAOKLWBEAPLEEM 600  
DB 541 SCWADYGAPVFPPLAIGGYKGDYSBAEALIMTFSLNNYPAGPRLAQAOKLWBEAPLEEM 600  
QY 601 RAPQRMAGHPQVFTFAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSWSRV 660  
DB 601 RAPQRMAGHPQVFTFAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSWSRV 660  
QY 661 MYDSKATLGLGGVAVVLGAVMAAGPFSYLGIRSSILVILQVVPFLVLSVGADNIFIFVLE 720  
DB 661 MYDSKATLGLGGVAVVLGAVMAAGPFSYLGIRSSILVILQVVPFLVLSVGADNIFIFVLE 720  
QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALICFFLGLALTPMPAVRTFALTSLGLAV 780  
DB 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALICFFLGLALTPMPAVRTFALTSLGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKROEASRLDYCCVKPQELPPPGQEGELLGFFQKAYAPFL 840  
DB 781 ILDFLLQMSAFVALLSLDSKROEASRLDYCCVKPQELPPPGQEGELLGFFQKAYAPFL 840  
QY 841 LHWITRGVLLILFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGVAPV 900  
DB 841 LHWITRGVLLILFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGVAPV 900  
QY 901 YFVTTLGYNFSSEAGNNAICSSAGCNPFSTQKIQVATEFPQSQYLAIIPASSWVDDFIDW 960  
DB 901 YFVTTLGYNFSSEAGNNAICSSAGCNPFSTQKIQVATEFPQSQYLAIIPASSWVDDFIDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVQFHKYLPWFNDLRP 1020  
DB 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVQFHKYLPWFNDLRP 1020  
QY 1021 NIKCPKGLAAYSTSNLTSDQVLAASRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
DB 1021 NIKCPKGLAAYSTSNLTSDQVLAASRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
QY 1081 RKVPGTDPAFEPVPTITNVFYEQYLTILPEGLFMLSCLCLVPTFAVSCLLGLDLSGGL 1140  
DB 1081 RKVPGTDPAFEPVPTITNVFYEQYLTILPEGLFMLSCLCLVPTFAVSCLLGLDLSGGL 1140  
QY 1141 NLLSVIMILVDTVGFMAWMDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
DB 1141 NLLSVIMILVDTVGFMAWMDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVAGVAMTNLPGIILVGLAKAQLIQIIPFFRNLNLTITLGLLHGLVFLPV 1260  
DB 1201 AKEATISMGSAVAGVAMTNLPGIILVGLAKAQLIQIIPFFRNLNLTITLGLLHGLVFLPV 1260  
QY 1261 ILSYVGPVNPALAEQKRAEAAVAVMVASCPNHPRSRVSTADNLYVNHSPSGSIKGAGA 1320  
DB 1261 ILSYVGPVNPALAEQKRAEAAVAVMVASCPNHPRSRVSTADNLYVNHSPSGSIKGAGA 1320  
QY 1321 ISNFLPNNGRQF 1332  
DB 1321 ISNFLPNNGRQF 1332  
RESULT 3  
AEB93570  
ID AEB93570 standard; protein; 1332 AA.  
XX  
AC AEB93570;  
XX  
DT 06-OCT-2005 (first entry)  
XX

Mon Apr 10 07:16:14 2006

Human NPC1L1 protein.

Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic; cardiant; vasotropic; hyperlipidemia; atherosclerosis; stroke; arteriosclerosis; human.

Homo sapiens.

WO2005069900-A2.

04-AUG-2005.

14-JAN-2005; 2005WO-US001469.

16-JAN-2004; 2004US-0537341P.

(MERI ) MERCK & CO INC.

Garcia-Calvo M;

WPI; 2005-564070/57.

N-PSDB; AEB93569.

Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by contacting NPC1L1 with detectably labeled substituted 2-azetidinone glucuronide and a candidate compound and determining if compound binds to human NPC1L1.

Example 2; SEQ ID NO 4; 215pp; English.

The invention relates to identifying a ligand of NPC1L1. The method involves contacting human NPC1L1 with a detectably labeled substituted 2-azetidinone glucuronide and a candidate compound and determining whether the candidate compound binds to human NPC1L1. In identifying a ligand of NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10 nM or lower. The detectably labeled substituted 2-azetidinone glucuronide is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands are useful for stimulating or blocking the activity of NPC1L1, and for treating conditions caused or mediated by NPC1L1. It is useful for reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart disease, stroke, or arteriosclerosis. The present sequence represents a human NPC1L1, a N-glycosylated protein.

Sequence 1332 AA;

Query Match	100.0%;	Score 6909;	DB 9;	Length 1332;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1332;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAEAGRLGWLWALLRLAQLSEPTTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN	60	
Db	1	MAEAGRLGWLWALLRLAQLSEPTTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN	60	
QY	61	TPARKITGDHLLILKQICPRLYTGPNQACCSAKQLVSLSEASITKALLTRCPACSDNF	120	
Db	61	TPARKITGDHLLILKQICPRLYTGPNQACCSAKQLVSLSEASITKALLTRCPACSDNF	120	
QY	121	VNLHCHNTCSNQSIFINVTVAQIGAGQLPAVVAEAFYQHSFAEQSYDCSRVRVPA	180	
Db	121	VNLHCHNTCSNQSIFINVTVAQIGAGQLPAVVAEAFYQHSFAEQSYDCSRVRVPA	180	
QY	181	ATLAVGTWGVYGSALCNAQRLNFPQGTGNGLAPLDITFHLLFPFGQAVSGIQLNEGV	240	
Db	181	ATLAVGTWGVYGSALCNAQRLNFPQGTGNGLAPLDITFHLLFPFGQAVSGIQLNEGV	240	
QY	241	ARCNESQGDVATCSCODCAACPAIARPOALDSTFYLGQPGSLVLIILICSVFAVVTI	300	
Db	241	ARCNESQGDVATCSCODCAACPAIARPOALDSTFYLGQPGSLVLIILICSVFAVVTI	300	
QY	301	LLVGFVRVARDKSMVDKGTSLSDKLSFSTHLLGQFGWGCTWVASWPLTILVLSV	360	
Db	301	LLVGFVRVARDKSMVDKGTSLSDKLSFSTHLLGQFGWGCTWVASWPLTILVLSV	360	

RESULT 4  
AAW79169  
ID AAW79169 standard; protein; 1332 AA.  
XX

QY	361	IPVVALAAGLVFTLTTDPVELWSAPNSQARSEKAFHQHFGFPFFRTNOVILTAPNRSSY	420	
Db	361	IPVVALAAGLVFTLTTDPVELWSAPNSQARSEKAFHQHFGFPFFRTNOVILTAPNRSSY	420	
QY	421	RYDSLILGPKNFSGILDLDLLELLELQERLRLHQLWSPQAQRNLSIQDICVAPLNPDMT	480	
Db	421	RYDSLILGPKNFSGILDLDLLELLELQERLRLHQLWSPQAQRNLSIQDICVAPLNPDMT	480	
QY	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTLAL	540	
Db	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTLAL	540	
QY	541	SCMADYGAPVFPFLAIGYKGYSEAEALIMTFSLNNYPAGDPRLAQAKLWEAELEEM	600	
Db	541	SCMADYGAPVFPFLAIGYKGYSEAEALIMTFSLNNYPAGDPRLAQAKLWEAELEEM	600	
QY	601	RAFQRRMAGMFQVTTAERSLEDEINRTAEDLPIFATSYIVIFLYISLALGSYSSSRV	660	
Db	601	RAFQRRMAGMFQVTTAERSLEDEINRTAEDLPIFATSYIVIFLYISLALGSYSSSRV	660	
QY	661	MYDSKATILGLGVAVVLGAVMAAMGFFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLE	720	
Db	661	MYDSKATILGLGVAVVLGAVMAAMGFFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLE	720	
QY	721	YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSEALCEFLGALTMPAVRTFALTSLGLAV	780	
Db	721	YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSEALCEFLGALTMPAVRTFALTSLGLAV	780	
QY	781	ILDFLLQMSAFVALLSLDSKQESRLDVCCCKVQPELPPQGGEGLLGFFQKAYAPFL	840	
Db	781	ILDFLLQMSAFVALLSLDSKQESRLDVCCCKVQPELPPQGGEGLLGFFQKAYAPFL	840	
QY	841	LHWITRGVLLFLALFGVLSYMSCHISVGLDQELALPKDSVLLDYFLNRYFVVGAPV	900	
Db	841	LHWITRGVLLFLALFGVLSYMSCHISVGLDQELALPKDSVLLDYFLNRYFVVGAPV	900	
QY	901	YFVTTLGYNFSSAGNNAICSSAGCANNFSTQKIQVATEFFPQSYLAIPASSWVDDFIDW	960	
Db	901	YFVTTLGYNFSSAGNNAICSSAGCANNFSTQKIQVATEFFPQSYLAIPASSWVDDFIDW	960	
QY	961	LTSSCCRLYISGPNKDKFCPTVNSLNCIKNCMSITMGSVRPSVQFHKYLPWFILNDRP	1020	
Db	961	LTSSCCRLYISGPNKDKFCPTVNSLNCIKNCMSITMGSVRPSVQFHKYLPWFILNDRP	1020	
QY	1021	NIKCPKGLAAYSTSVNLTSDGVLSRFAHYHKLKNSODYTEALRAARELAANITADL	1080	
Db	1021	NIKCPKGLAAYSTSVNLTSDGVLSRFAHYHKLKNSODYTEALRAARELAANITADL	1080	
QY	1081	RKVPGTDPAPFVFPYTTTNNVFYEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSGLL	1140	
Db	1081	RKVPGTDPAPFVFPYTTTNNVFYEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSGLL	1140	
QY	1141	NLLSIWMLVDVTGFMALWDISYNVAVSLINLVSAVGMVSFVSHITSFSAISTKPTWLER	1200	
Db	1141	NLLSIWMLVDVTGFMALWDISYNVAVSLINLVSAVGMVSFVSHITSFSAISTKPTWLER	1200	
QY	1201	AKEATISMGSAVEAGVAMTNLPGIILVLGLAKAQILQIFFPRLNLLITLGLLHGLVFLPV	1260	
Db	1201	AKEATISMGSAVEAGVAMTNLPGIILVLGLAKAQILQIFFPRLNLLITLGLLHGLVFLPV	1260	
QY	1261	ILSYVGPDPNVPALAEQKRAEAAVAVMVASCPNHPSPVSTADNIYVNHSPFSGIKAGA	1320	
Db	1261	ILSYVGPDPNVPALAEQKRAEAAVAVMVASCPNHPSPVSTADNIYVNHSPFSGIKAGA	1320	
QY	1321	ISNFLPNNGRQF 1332		
Db	1321	ISNFLPNNGRQF 1332		



AC	AAW79169;	
XX	06-NOV-2001 (first entry)	
DT	Human protein SEQ ID NO 1831.	
DE		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorder; arthritis; inflammation.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200157190-A2.	
XX		
PD	09-AUG-2001.	
XX		
PP	05-FEB-2001; 2001WO-US004098.	
XX		
PR	03-FEB-2000; 2000US-00496914.	
PR	27-APR-2000; 2000US-00560875.	
PR	20-JUN-2000; 2000US-00598075.	
PR	19-JUL-2000; 2000US-00620325.	
PR	01-SEP-2000; 2000US-00654936.	
PR	15-SEP-2000; 2000US-00663561.	
PR	20-OCT-2000; 2000US-00693325.	
PR	30-NOV-2000; 2000US-00728422.	
XX	(HYSE-) HYSEQ INC.	
PA		
XX		
PI	Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;	
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Chen ZW;	
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;	
XX		
DR	WPI: 2001-476283/51.	
DR	N-PSDB; AAK52302.	
XX		
PT	Nucleic acids encoding polypeptides with cytokine-like activities, useful	
PT	in diagnosis and gene therapy.	
XX		
PS	Claim 20; Page 4216-4219; 6221pp; English.	
XX		
CC	The invention relates to polynucleotides (AAK51456-AAK53435) and the	
CC	encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to	
CC	cytokine, cell proliferation or cell differentiation or which may induce	
CC	production of other cytokines in other cell populations. The	
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or	
CC	peptide therapy. The polypeptides have various cytokine-like activities,	
CC	e.g. stem cell growth factor activity, haematopoiesis regulating	
CC	activity, tissue growth factor activity, immunomodulatory activity and	
CC	activin/inhibin activity and may be useful in the diagnosis and/or	
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and	
CC	inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111	
CC	(AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the	
CC	sequence listing were missing at the time of publication	
XX		
SQ	Sequence 1332 AA;	
	Query Match 99.8%; Score 6896; DB 4; Length 1332;	
	Best Local Similarity 99.8%; Pred. No. 0;	
	Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
QY	1 MAEAGLGRWLLWALLRLAQSEPYTTIHQPGYCAFYDECGKPNELSGSLMTLSNVCSLSN 60	
DB		
DB	1 MAEAGLGRWLLWALLRLAQSEPYTTIHQPGYCAFYDECGKPNELSGSLMTLSNVCSLSN 60	
QY	61 TPARKITGDHLLLOKICPRLVTGNTQACCSAKQLVSLASLSITKALLTRCPACSDNF 120	
DB		
DB	61 TPARKITGDHLLLOKICPRLVTGNTQACCSAKQLVSLASLSITKALLTRCPACSDNF 120	
QY	121 VNLHCHNTCSNQSIFINVTVAQLGAGQLPAVAYEAFYQHSFAEQSYDSCSRVRVPA 180	
DB		
DB	121 VNLHCHNTCSNQSIFINVTVAQLGAGQLPAVAYEAFYQHSFAEQSYDSCSRVRVPA 180	

QY	181 ATLAVGTMCGVYGSALCNAQRWLNFGQDGTGNGIAPLDITFHLLEPQAVGSGIQPLNEGV 240	
DB		
DB	181 ATLAVGTMCGVYGSALCNAQRWLNFGQDGTGNGIAPLDITFHLLEPQAVGSGIQPLNEGV 240	
QY	241 ARCNSQGGDVATCSCODCAACPAIARPOALDSTFYLGOMPGSLVLIILICSVFAVVTI 300	
DB		
DB	241 ARCNSQGGDVATCSCODCAACPAIARPOALDSTFYLGOMPGSLVLIILICSVFAVVTI 300	
QY	301 LLVGRFVAPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTILVLSV 360	
DB		
DB	301 LLVGRFVAPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTILVLSV 360	
QY	361 IPVVALAAGLVFTELTTPDVELMSAPNSQARSEKAFHDQHGFPFRFTNQVILTPNRSSY 420	
DB		
DB	361 IPVVALAAGLVFTELTTPDVELMSAPNSQARSEKAFHDQHGFPFRFTNQVILTPNRSSY 420	
QY	421 RYDSLILGPKNFGILDLDLLELELQERLHLQVMSPEAQRNLSLODICYAPLNPDT 480	
DB		
DB	421 RYDSLILGPKNFGILDLDLLELELQERLHLQVMSPEAQRNLSLODICYAPLNPDT 480	
QY	481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOVDWKDHFLYCANAPLTFKDGTFALAL 540	
DB		
DB	481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOVDWKDHFLYCANAPLTFKDGTFALAL 540	
QY	541 SCMADYGAPVPPFLAIGGYKGKDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600	
DB		
DB	541 SCMADYGAPVPPFLAIGGYKGKDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600	
QY	601 RAFQRMAGMGOVTFEATERSLEDEINRTTAEDLPIFATSYIVIFLISIALGYSYSSWSRV 660	
DB		
DB	601 RAFQRMAGMGOVTFEATERSLEDEINRTTAEDLPIFATSYIVIFLISIALGYSYSSWSRV 660	
QY	661 MYDSKATILGLGGAVVVLGAVMAAMGFFSYLGRSSILVILQVVPFLVLSVGADNIFIFVLE 720	
DB		
DB	661 MYDSKATILGLGGAVVVLGAVMAAMGFFSYLGRSSILVILQVVPFLVLSVGADNIFIFVLE 720	
QY	721 YQRLPRPCEPREVHIGRALGRVAPSMLLCSLSEACFFLGALTMPAVRTFALTSGLAV 780	
DB		
DB	721 YQRLPRPCEPREVHIGRALGRVAPSMLLCSLSEACFFLGALTMPAVRTFALTSGLAV 780	
QY	781 ILDFLLQMSAFVALLSLDSKQESRLDVCCCKVQKQELPPQCGEGLLIGFFQKAYAPFL 840	
DB		
DB	781 ILDFLLQMSAFVALLSLDSKQESRLDVCCCKVQKQELPPQCGEGLLIGFFQKAYAPFL 840	
QY	841 LHWITRGVVLFLALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGPV 900	
DB		
DB	841 LHWITRGVVLFLALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGPV 900	
QY	901 YFVTTLGYNFSSEAGMNAICSSAGCENFSFTQKIQYATEFFPQSYLAIPASSWVDDFIDW 960	
DB		
DB	901 YFVTTLGYNFSSEAGMNAICSSAGCENFSFTQKIQYATEFFPQSYLAIPASSWVDDFIDW 960	
QY	961 LTPSSCCRLYISGPNKDKFCPSVTNSLNCNMSITMGSVRPSVEQHKYLPWFINDRP 1020	
DB		
DB	961 LTPSSCCRLYISGPNKDKFCPSVTNSLNCNMSITMGSVRPSVEQHKYLPWFINDRP 1020	
QY	1021 NIKCPKGLAAYSTSVNLTSDQVLAASFMAVHKPKNSQDYTEALRAARELAANTADL 1080	
DB		
DB	1021 NIKCPKGLAAYSTSVNLTSDQVLAASFMAVHKPKNSQDYTEALRAARELAANTADL 1080	
QY	1081 RKVPGTDPAFVFPYTTITNVFVEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSRSGLL 1140	
DB		
DB	1081 RKVPGTDPAFVFPYTTITNVFVEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLSRSGLL 1140	
QY	1141 NLLSVIMLVDTVGFMAWDISYNVAVSLINLVASGVMSVFVSHITRSFAISTKTWLER 1200	
DB		
DB	1141 NLLSVIMLVDTVGFMAWDISYNVAVSLINLVASGVMSVFVSHITRSFAISTKTWLER 1200	
QY	1201 AKEATISMGSAVFAVAMTNLPGILVLGLAKAQLIQIFFFRNLNLLITLGLLHGVFLPV 1260	
DB		
DB	1201 AKEATISMGSAVFAVAMTNLPGILVLGLAKAQLIQIFFFRNLNLLITLGLLHGVFLPV 1260	

Qy	1261	ILSYVGP	DVNPALALEOKRAEEAAVAVMVASCPNHPRSRVSTADNIYVNHSPFEGSIKGAGA	1320
Db	1261	ILSYVGP	DVNPALALEOKRAEEAAVAVMVASCPNHPRSRVSTADNIYVNHSPFEGSIKGAGA	1320
Qy	1321	ISNFLPNNGRQF	1332	
Db	1321	ISNFLPNNGRQF	1332	
RESULT 5				
ID	AAG65638	AAG65638 standard; protein; 1332 AA.		
XX	AC	AAG65638;		
XX	DT	07-JAN-2002 (first entry)		
XX	DE	SSD-containing SSP1 protein sequence.		
XX	KW	SSD; sterol-sensing domain; human; liver; testis; brain; cancer;		
XX	KW	nootropic; neuroprotective; antidiabetic; antiarteriosclerotic;		
XX	KW	cytostatic; antilipemic; SSPI.		
XX	OS	Homo sapiens.		
XX	PN	WO200170974-A1.		
XX	PD	27-SEP-2001.		
XX	PF	22-MAR-2001; 2001WO-JP002279.		
XX	PR	24-MAR-2000; 2000JP-00088595.		
XX	PA	(TAKE ) TAKEDA CHEM IND LTD.		
XX	PI	Taniyama Y, Kita S, Komiyama T;		
XX	DR	WPI; 2001-611501/70.		
XX	DR	N-PSDB; AAI66914.		
XX	PT	New sterol-sensing domain-containing protein for diagnosing and screening		
XX	PT	candidate compounds in drug development for diabetes, obesity, cancer,		
XX	PT	arteriosclerosis, hyperlipidemia and neurodegenerative disorders.		
XX	PS	Claim 7; Page 115-122; 171pp; Japanese.		
XX	CC	The invention provides a novel SSD (sterol-sensing domain)-containing		
XX	CC	protein. The protein originates from human liver, human testis or human		
XX	CC	brain. The protein can be expressed by standard recombinant methodology.		
XX	CC	The proteins, encoded DNAs and antibodies are useful in diagnosis and		
XX	CC	screening candidate compounds in drug development for diabetes, obesity,		
XX	CC	cancer, arteriosclerosis, hyperlipemia, neurodegenerative disorders such		
XX	CC	as Alzheimer's disease and neural disorders. The present sequence		
XX	CC	represents the SSPI protein which contains the SSD domain		
XX	SSQ	Sequence 1332 AA;		
Query Match 99.8%; Score 6896; DB 4; Length 1332;				
Best Local Similarity 99.8%; Pred. No. 0;				
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
Qy	1	MAEAGLRCGLWALLRLAQSEPYTHIQPGVCAPYDECGKNPELSGSLMTLSNVCSLSN	60	
Db	1	MAEAGLRCGLWALLRLAQSEPYTHIQPGVCAPYDECGKNPELSGSLMTLSNVCSLSN	60	
Qy	61	TPARKITGDHLILLOKICPRLYTGNTQACCSAKQLVSLASLSITKALLTRCPACSDNF	120	
Db	61	TPARKITGDHLILLOKICPRLYTGNTQACCSAKQLVSLASLSITKALLTRCPACSDNF	120	
Qy	121	VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFQHSFABQSYDSCSRVRPAA	180	
Db	121	VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFQHSFABQSYDSCSRVRPAA	180	

Qy	181	ATLAVGTMCVYGSALCNAQRWLNFOGDTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV	240
Db	181	ATLAVGTMCVYGSALCNAQRWLNFOGDTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV	240
Qy	241	ARCNEQGGDDVATCSQDCAASCPAIARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI	300
Db	241	ARCNEQGGDDVATCSQDCAASCPAIARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI	300
Qy	301	LLVGFVRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASWPLTILVLSV	360
Db	301	LLVGFVRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASWPLTILVLSV	360
Qy	361	IPVVALAAGLVFTFELTTDPVELWSAPNSQARSEKAFHDQHFQFFFRTNQVILTAPNRSSY	420
Db	361	IPVVALAAGLVFTFELTTDPVELWSAPNSQARSEKAFHDQHFQFFFRTNQVILTAPNRSSY	420
Qy	421	RYDSLILGPKNFGSILDLDLLLELLELQERLRLQVWSPQAQRNISLODICYAPLNPDNT	480
Db	421	RYDSLILGPKNFGSILDLDLLLELLELQERLRLQVWSPQAQRNISLODICYAPLNPDNT	480
Qy	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOVDWKDHFLYCANAPLTFKDGOTALAL	540
Db	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGOTSOVDWKDHFLYCANAPLTFKDGOTALAL	540
Qy	541	SCMADYGAPVPFPLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLWEEAFLEEM	600
Db	541	SCMADYGAPVPFPLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLWEEAFLEEM	600
Qy	601	RAFQRMAGMFQVTTAERSLEDEINRTTAEDLPFIFATSYIVIFLYISLALGYSYSSWSRV	660
Db	601	RAFQRMAGMFQVTTAERSLEDEINRTTAEDLPFIFATSYIVIFLYISLALGYSYSSWSRV	660
Qy	661	MVDSKATLGLGGVAVVLGAVMAAGFFSYLIGSSILVQVVPFLVLSVGADNIFIFVLE	720
Db	661	MVDSKATLGLGGVAVVLGAVMAAGFFSYLIGSSILVQVVPFLVLSVGADNIFIFVLE	720
Qy	721	YQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSEACFFLGTALTPMPAVRTFALTSLGLAV	780
Db	721	YQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSEACFFLGTALTPMPAVRTFALTSLGLAV	780
Qy	781	ILDFLOMSAFVALLSLDSKROBASRLDVCCKVKPQBLPPPGQEGLLLGFFQKAYAPFL	840
Db	781	ILDFLOMSAFVALLSLDSKROBASRLDVCCKVKPQBLPPPGQEGLLLGFFQKAYAPFL	840
Qy	841	LHWITRGVLLLLFLALFGVSLYSNMCHTISVGLDQSLALPKDSYLLDYFLFLNRYFEVQAPV	900
Db	841	LHWITRGVLLLLFLALFGVSLYSNMCHTISVGLDQSLALPKDSYLLDYFLFLNRYFEVQAPV	900
Qy	901	YFVTTLGYNFSSEAGMNAICSSAGCANNFSTQKIQYATEPPEQSYLAIPASSWVDDFDWM	960
Db	901	YFVTTLGYNFSSEAGMNAICSSAGCANNFSTQKIQYATEPPEQSYLAIPASSWVDDFDWM	960
Qy	961	LTSSCCRLYISGPNKDKFCPTVNSLNCIKNCMSITMGSVRPSVEQFHKYLPWFLNDRP	1020
Db	961	LTSSCCRLYISGPNKDKFCPTVNSLNCIKNCMSITMGSVRPSVEQFHKYLPWFLNDRP	1020
Qy	1021	NICKCPKGLAAYSTVNLTSDDGVLASRFMAYHKPLKNSODYTEALRAARELANITADL	1080
Db	1021	NICKCPKGLAAYSTVNLTSDDGVLASRFMAYHKPLKNSODYTEALRAARELANITADL	1080
Qy	1081	RKVEGTDPAFEVFPYTTITNVFYEQYLTILPEGLPMLSLCLVPTFAVSCLLGLDLRSGLL	1140
Db	1081	RKVEGTDPAFEVFPYTTITNVFYEQYLTILPEGLPMLSLCLVPTFAVSCLLGLDLRSGLL	1140
Qy	1141	NLLSIWMLVDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPALSTKPTWLER	1200
Db	1141	NLLSIWMLVDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPALSTKPTWLER	1200
Qy	1201	AKEATIISWGSAPVAGVAMTNLPGILVLGLAKAQILQIFFFFRLNLLITLLGLLHGLVFLPV	1260
Db	1201	AKEATIISWGSAPVAGVAMTNLPGILVLGLAKAQILQIFFFFRLNLLITLLGLLHGLVFLPV	1260
Qy	1261	ILSVYVDPVNPALALEOKRAEEAAVAVMVASCPNHPRSRVSTADNIYVNHSPFEGSIKGAGA	1320

|||||  
Db 1261 ILSYVGPDPNPALEQKRAEAAVMVASCPNHPSTADNIYVNHSPFSGSIKAGA 1320  
QY 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332  
RESULT 6  
AAM79168  
ID AAM79168 standard; protein; 1359 AA.  
AC AAM79168;  
XX  
XX 06-NOV-2001 (first entry)  
XX Human protein SEQ ID NO 1830.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200157190-A2.  
PN  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US004098.  
XX  
XX 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
PR 20-JUN-2000; 2000US-00598075.  
PR 19-JUL-2000; 2000US-00620325.  
PR 01-SEP-2000; 2000US-00654936.  
PR 15-SEP-2000; 2000US-00663561.  
PR 20-OCT-2000; 2000US-00693325.  
PR 30-NOV-2000; 2000US-00728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
PI Ma Y, Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52301.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful  
PT in diagnosis and gene therapy.  
PT  
XX  
XX Claim 20; Page 4213-4216; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK53581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
XX Sequence 1359 AA;

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 MAEAGLRGWLILWALLRLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCLSN 60  
Db 1 MAEAGLRGWLILWALLRLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCLSN 60  
QY 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASLSITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASLSITKALLTRCPACSDNF 120  
QY 121 VNLHCHNTCSNQSIFINVTRVAQLGAGOLPAVAYEAFYQHSFABQSVDSCSRVPAA 180  
Db 121 VNLHCHNTCSNQSIFINVTRVAQLGAGOLPAVAYEAFYQHSFABQSVDSCSRVPAA 180  
QY 181 ATLAVGTMCGVYGVSALCNAQRWLNFGQDTGNGLAFLDITFHLLPEQQAAGVSGIQLNEG 240  
Db 181 ATLAVGTMCGVYGVSALCNAQRWLNFGQDTGNGLAFLDITFHLLPEQQAAGVSGIQLNEG 240  
QY 241 ARCNSQGDVATCSCQDCAASCPAIAAPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300  
Db 241 ARCNSQGDVATCSCQDCAASCPAIAAPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300  
QY 301 LLVGFRVAPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVLSV 360  
Db 301 LLVGFRVAPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWASWPLTILVLSV 360  
QY 361 IPVVALAAGLVFTLTTPVELWNSAPNSOARSEKAFHDOHFGPFFRTNQVILTAPNRSSY 420  
Db 361 IPVVALAAGLVFTLTTPVELWNSAPNSOARSEKAFHDOHFGPFFRTNQVILTAPNRSSY 420  
QY 421 RYDSLILGPKNFGSGLDLDLLELELERLHLQVWSPQAQRNLSLODICYAPLNPDNT 480  
Db 421 RYDSLILGPKNFGSGLDLDLLELELERLHLQVWSPQAQRNLSLODICYAPLNPDNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKHDFLYCANAPLTFKDGTALAL 540  
Db 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKHDFLYCANAPLTFKDGTALAL 540  
QY 541 SCMADYGAPVFPFLAIGGYKGDYSEAEALMTFSLNNYPAGDPRLAOKLWEAEFLBEM 600  
Db 541 SCMADYGAPVFPFLAIGGYKGDYSEAEALMTFSLNNYPAGDPRLAOKLWEAEFLBEM 600  
QY 601 RAPQRMAGMQVFTTAERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSYSSWSRV 660  
Db 601 RAPQRMAGMQVFTTAERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSYSSWSRV 660  
QY 661 MVDSKATLGLGGVAVVLGAVMAAGFFSYLGRSLVILQVVPFLVSVGADNIIFFVLE 720  
Db 661 MVDSKATLGLGGVAVVLGAVMAAGFFSYLGRSLVILQVVPFLVSVGADNIIFFVLE 720  
QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSAICPFLGALTMPMPAVRTFALTSLGLAV 780  
Db 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSAICPFLGALTMPMPAVRTFALTSLGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKQEAASRLDVCCCVKPEQLPPPGQEGEGLLGFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQEAASRLDVCCCVKPEQLPPPGQEGEGLLGFQKAYAPFL 840  
QY 841 LHWITRGVLLLFALFGVSLYSCHISVGLDQELALPKDSYILDYFLFLNRYFEVGPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSCHISVGLDQELALPKDSYILDYFLFLNRYFEVGPV 900  
QY 901 YFVVTTLGYNFSSAGMAICSSACGNFSPFKIQYATEPPEQSYLAIPASSWDDFTDW 960  
Db 901 YFVVTTLGYNFSSAGMAICSSACGNFSPFKIQYATEPPEQSYLAIPASSWDDFTDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPSVTNSLNCNCSITMGSVRPSVEQFHKLWPLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSVTNSLNCNCSITMGSVRPSVEQFHKLWPLNDRP 1020  
QY 1021 NIKCPKGLAAYSTSVNLTSDGQVL-----ASRFMAYH 1053  
Db 1021 NIKCPKGLAAYSTSVNLTSDGQVL-----ASRFMAYH 1053

QY 1054 KPLKNSQDYTEALRAARELANITADLRKVPGTDPAPFVFPYTTITNVFYEQYLTILPEGL 1113  
Db |||||||  
QY 1081 KPLKNSQDYTEALRAARELANITADLRKVPGTDPAPFVFPYTTITNVFYEQYLTILPEGL 1140  
Db |||||||  
QY 1114 FMLSCLVPTFAVSCLLGLDLRSGLNLSIVMILVDTVGFMAWDISYNAVSLINLVS 1173  
Db |||||||  
QY 1141 FMLSCLVPTFAVSCLLGLDLRSGLNLSIVMILVDTVGFMAWDISYNAVSLINLVS 1200  
QY 1174 AVGMSVEFVSHITRFAISTKPTWLERAKEATISMGSAVAGVAMTNLPGILVLGLAKAQ 1233  
Db |||||||  
QY 1201 AVGMSVEFVSHITRFAISTKPTWLERAKEATISMGSAVAGVAMTNLPGILVLGLAKAQ 1260  
QY 1234 LIQIFPRLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALQKRAEAEVAAVMVASCP 1293  
Db |||||||  
QY 1261 LIQIFPRLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALQKRAEAEVAAVMVASCP 1320  
QY 1294 NHPSRVSTADNIYVNHSEFGSIKAGAGISNPLPNNGRQF 1332  
Db |||||||  
QY 1321 NHPSRVSTADNIYVNHSEFGSIKAGAGISNPLPNNGRQF 1359

RESULT 7  
ADJ27337  
ID ADJ27337 standard; protein; 1359 AA.

AC ADJ27337;  
DT 20-MAY-2004 (first entry)  
XX Human NPC1L1.  
XX Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;  
KW plasma membrane; transport signal; promoter; SREBP1;  
KW binding regulated element binding protein 1; SREBP1;  
KW sterol regulated element binding protein 1; SREBP1;  
KW binding regulated element binding protein 1; SREBP1;  
KW SCD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;  
KW cholesterol absorption; serum cholesterol; hyperlipidaemia;  
KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.  
XX Homo sapiens.  
XX WO2004009772-A2.  
XX 29-JAN-2004.  
XX 17-JUL-2003; 2003WO-US022467.  
XX 19-JUL-2002; 2002US-0397442P.  
XX (SCHE ) SCHERING CORP.  
XX Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;  
XX WPI; 2004-132945/13.  
XX N-PSDB; ADJ27336.  
XX GENBANK; AF192522.  
XX New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting  
PT or identifying antagonists of NPC1L1 for inhibiting intestinal  
PT cholesterol absorption in a subject, or for treating elevated serum  
PT cholesterol or stroke.  
XX Example 7; SEQ ID NO 44; 125pp; English.  
XX This sequence represents a Niemann-Pick disease, type C1, gene-like 1  
CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains  
CC a motif ADJ27331 which acts as a trans-golgi network to plasma membrane  
CC transport signal, and which exhibits limited tissue distribution and  
CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a  
CC sterol regulated element binding protein 1 (SREBP1) binding consensus  
CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-  
CC sensing domain (SSD) which is involved in sensing cholesterol levels,  
CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1

CC has 42% amino acid sequence homology to human NPC1, a receptor  
CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or  
CC polynucleotides are useful for detecting or identifying antagonists of  
CC NPC1L1, which can be used for inhibiting intestinal cholesterol  
CC absorption in a subject, or for treating medical conditions including  
CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary  
CC heart disease, stroke or arteriosclerosis.  
XX Sequence 1359 AA;  
SQ  
Query Match 99.5%; Score 6872.5; DB 8; Length 1359;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;  
QY 1 MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKNPELSGSLMTLSNVCLSN 60  
Db |||||||  
QY 1 MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKNPELSGSLMTLSNVCLSN 60  
QY 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Db |||||||  
QY 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
QY 121 VNLHCHTCSPNQSLFINVTRVAOLGAGOLPAVAYEAFYQHSFAEQSDSCSRVPA 180  
Db |||||||  
QY 121 VNLHCHTCSPNQSLFINVTRVAOLGAGOLPAVAYEAFYQHSFAEQSDSCSRVPA 180  
QY 181 ATLAVGTCGVTGSGALCNQRLNFOGDTGNGLAPLDITFHLLEPGQAVGSGIQLNEGV 240  
Db |||||||  
QY 181 ATLAVGTCGVTGSGALCNQRLNFOGDTGNGLAPLDITFHLLEPGQAVGSGIQLNEGV 240  
QY 241 ARCNEQSGDDVATCSQDCAASPATARPQALDSTFYLQMPGSLVLIILCSVFAVVTI 300  
Db |||||||  
QY 301 LLVGFVAPARDKSMVDPKGTSLSDKLSFSFTHLLGPOFGWGTWVASWPLTILVLSV 360  
Db |||||||  
QY 361 IPVVALAAGLVFTFELTTDPVELWSAPNSQARSEKAFHDQHFQFFRTNQVILTPNRSSY 420  
Db |||||||  
QY 361 IPVVALAAGLVFTFELTTDPVELWSAPNSQARSEKAFHDQHFQFFRTNQVILTPNRSSY 420  
QY 421 RYDSSLGPNFSGIILDLLELLELQERLRLHQLVMSPEAQRNISLQDI CYAPLNPDT 480  
Db |||||||  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGITAL 540  
Db |||||||  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGITAL 540  
QY 541 SCMDYGA PVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600  
Db |||||||  
QY 541 SCMDYGA PVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600  
QY 601 RAFORMMAGFQVTFATRSLEDEINRTAEDLPFATSYIVIFLYISLALGYSWSRV 660  
Db |||||||  
QY 661 MVDKATILGLGVAVNLGAVMAAGFYSYLGIBSSVLVQVPLVLSVGDADNIFIVLE 720  
Db |||||||  
QY 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSEAICFFLGALTPMPAVRTFALTSGLAV 780  
Db |||||||  
QY 781 ILDFLLQMSAFVALLSDSRQASRLDVCCCKVPQBLPPPGQEGLLGLFFQKAYAPFL 840  
Db |||||||  
QY 841 LHMWITRGVLLFLALFGVSLYSMCHTSVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900  
Db |||||||

QY 901 YFVTTTGYNFSSAGNAICSSAGCANNFSPTQKIQVATEPPEQSYLAIPASSWVDDFDW 960  
DB 901 YFVTTTGYNFSSAGNAICSSAGCANNFSPTQKIQVATEPPEQSYLAIPASSWVDDFDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCUKNCMSITMGSVRSVQFHKYLPWFNLDRP 1020  
DB 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCUKNCMSITMGSVRSVQFHKYLPWFNLDRP 1020  
QY 1021 NIKCPKGLAAYSTVNLSDGQVL-----ASRPMAYH 1053  
DB 1021 NIKCPKGLAAYSTVNLSDGQVLDVVALSPRLEYSGTISAHCNLYLLDLSASRRPMAYH 1080  
QY 1054 KPLKNSQDYTEALRAARELAANITADLRKVPDTPAFEPVPPYTTTNVFEQYLTILPEGL 1113  
DB 1081 KPLKNSQDYTEALRAARELAANITADLRKVPDTPAFEPVPPYTTTNVFEQYLTILPEGL 1140  
QY 1114 FMLSCLVPTFAVSCILLGLDLSGLNLLSIYVILVDTVGFNALWDISNAVSLNLVYS 1173  
DB 1141 FMLSCLVPTFAVSCILLGLDLSGLNLLSIYVILVDTVGFNALWGISNAVSLNLVYS 1200  
QY 1174 AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1233  
DB 1201 AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1260  
QY 1234 LIQIFPFRNLTLTLGLLHGLVFLPVLVLSYVGPVDPNPALEQKRAEBAVAAMVVASCP 1293  
DB 1261 LIQIFPFRNLTLTLGLLHGLVFLPVLVLSYVGPVDPNPALEQKRAEBAVAAMVVASCP 1320  
QY 1294 NHPSRVSTADNIYVNHSPFGSIKGAGAIISNPLNNGROF 1332  
DB 1321 NHPSRVSTADNIYVNHSPFGSIKGAGAIISNPLNNGROF 1359

RESULT 8

ADY60780  
ID ADY60780 standard; protein; 1359 AA.  
AC ADY60780;  
XX  
DT 19-MAY-2005 (first entry)  
DE Human Niemann-Pick C1-like protein 1 (NPC1L1) protein SEQ ID 44.  
XX  
KW Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;  
KW Cardiant; Vasotropic; Cerebroprotective; cholesterol; hyperlipidemia;  
KW metabolic disorder; atherosclerosis; cardiovascular disease;  
KW coronary artery disease; cerebrovascular ischemia; neurological disease;  
KW arteriosclerosis; transgenic animal.  
OS Homo sapiens.  
XX  
PN WO2005015988-A1.  
XX  
PD 24-FEB-2005.  
XX  
PF 16-DEC-2003; 2003WO-US040113.  
XX  
PR 17-JUL-2003; 2003US-00621758.  
PR 22-AUG-2003; 2003US-00646301.  
PR 16-SEP-2003; 2003US-00663208.  
XX  
PA (SCHE ) SCHERING CORP.  
XX  
PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;  
XX WPI; 2005-284403/29.  
DR N-P8DB; ADY60779.  
DR  
XX  
PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of  
PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for  
PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.  
XX

Disclosure; SEQ ID NO 44; 146pp; English.

PS The invention relates to an isolated Niemann-Pick C1-like protein 1  
XX (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an  
CC amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,  
CC human and mouse NPC1L1). Also included are an isolated polynucleotide  
CC encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a  
CC recombinant vector comprising the polynucleotide, a host cell comprising  
CC the vector, an isolated antibody which specifically binds to NPC1L1 (or  
CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),  
CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous  
CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not  
CC produce any functional NPC1L1 protein), an offspring or progeny of the  
CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a  
CC sample for an intestinal sterol or Salpha-sterol absorption antagonist,  
CC inhibiting NPC1L1 mediated sterol or Salpha-sterol uptake in a subject, a  
CC kit (comprising ezetimibe or substituted azetidinone in a pharmaceutical  
CC dosage form, and information indicating that NPC1L1 is a target of  
CC ezetimibe or substituted azetidinone), decreasing the level of intestinal  
CC sterol or Salpha-sterol absorption in a subject (involving reducing the  
CC level of expression of NPC1L1 in the subject), identifying an antagonist  
CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a  
CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist  
CC of NPC1L1 which is useful for inhibiting or decreasing the level of  
CC NPC1L1 mediated sterol or Salpha-sterol uptake in a subject such as  
CC mouse, rat or human. The level of expression of NPC1L1 in the subject is  
CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is  
CC useful for screening a sample for intestinal sterol or Salpha-sterol  
CC absorption antagonist. The NPC1L1 antagonists identified are useful for  
CC treating medical conditions caused or mediated by NPC1L1, e.g.,  
CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or  
CC arteriosclerosis. The present sequence represents a mammalian NPC1L1  
XX protein.  
SQ Sequence 1359 AA;

Query Match 99.5%; Score 6872.5; DB 9; Length 1359;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;  
QY 1 MAEAGIRGWLWALLRLLAQSEPYTHIHPGCAFYDECKNPELSGSLMTLSNVCLSN 60  
DB 1 MAEAGIRGWLWALLRLLAQSEPYTHIHPGCAFYDECKNPELSGSLMTLSNVCLSN 60  
QY 61 TPARKITGDHLLILQKICPRLVTGNTOACCSAKOLVSEASLSITKALLTRCPACSDNF 120  
DB 61 TPARKITGDHLLILQKICPRLVTGNTOACCSAKOLVSEASLSITKALLTRCPACSDNF 120  
QY 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAEAFYQHSFAEQSYDSCSRVRPAA 180  
DB 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAEAFYQHSFAEQSYDSCSRVRPAA 180  
QY 181 ATLAVGTWCVYGSALCNAQRWLNFGDTGNGLAFLDITFHLLEPGQAVSGIQPLNEGV 240  
DB 181 ATLAVGTWCVYGSALCNAQRWLNFGDTGNGLAFLDITFHLLEPGQAVSGIQPLNEGV 240  
QY 241 ARCNEQGDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIIILCSVFAVVTI 300  
DB 241 ARCNEQGDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIIILCSVFAVVTI 300  
QY 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWVASWPLTIIVLSV 360  
DB 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWVASWPLTIIVLSV 360  
QY 361 IPVVALAAGLVPTETTTDPVELWSAPNSQARSEKAFHQHFGPPFRTRNQVILTAPNRSY 420  
DB 361 IPVVALAAGLVPTETTTDPVELWSAPNSQARSEKAFHQHFGPPFRTRNQVILTAPNRSY 420  
QY 421 RYDSLILGPKNFSGLTDLDDLLLELELOERLHRHQVWSPAQORNISLODICYAPLNPDNT 480  
DB 421 RYDSLILGPKNFSGLTDLDDLLLELELOERLHRHQVWSPAQORNISLODICYAPLNPDNT 480  
QY 481 SLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHPFLYCANAPLTFKDGATAL 540

Db 481 SLVDCINSLLQVFNRTLLLTANQTLMGQTSQVDKDFLYCANAPLTFKDGATLAL 540  
Qy 541 SCHADYCAPVFPFLAIGGYGKDYSEAEALIMTFSLNNYPAGDRLAQAKLWEAEFL 600  
Db 541 SCHADYCAPVFPFLAIGGYGKDYSEAEALIMTFSLNNYPAGDRLAQAKLWEAEFL 600  
Qy 601 RAFORMMAGMFQVTFTEAERSLEDEINRTAEDLPIFATSVIVIFLYISLALGSYSWSRV 660  
Db 601 RAFORMMAGMFQVTFTEAERSLEDEINRTAEDLPIFATSVIVIFLYISLALGSYSWSRV 660  
Qy 661 MVDKSKATLGLGVAVVLGAVMAAGFFSYLGIRSSVLQVVPFLVLSVGADNIFIVLE 720  
Db 661 MVDKSKATLGLGVAVVLGAVMAAGFFSYLGIRSSVLQVVPFLVLSVGADNIFIVLE 720  
Qy 721 YQRLPRPGPREVHIHGRALGRVAPSMMLCSLSEAI CFFLGALTPMPAVTFEALTSLAV 780  
Db 721 YQRLPRPGPREVHIHGRALGRVAPSMMLCSLSEAI CFFLGALTPMPAVTFEALTSLAV 780  
Qy 781 ILDFLLQMSAFVALLSDSKRQASRLDCCVKPQELPPPGQEGELLLGFFOKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSDSKRQASRLDCCVKPQELPPPGQEGELLLGFFOKAYAPFL 840  
Qy 841 LHMTRGVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHMTRGVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Qy 901 YFVTTLYNFSSEAGMNAICSSAGCNFSTQKIQYATEPPEQSYLAIPASSWDDFIDW 960  
Db 901 YFVTTLYNFSSEAGMNAICSSAGCNFSTQKIQYATEPPEQSYLAIPASSWDDFIDW 960  
Qy 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSQGVL-----ASRFWAYH 1053  
Db 1021 NIKCPKGLAAYSTSVNLTSQGVLDTVAILSPRLEYSGTISAHCNLYLLDSASRFWAYH 1080  
Qy 1054 KPLKNSQDYTEALRAARELANITADLRKVPDPAPEVPPYTTINVFYEQYLTIPEGL 1113  
Db 1081 KPLKNSQDYTEALRAARELANITADLRKVPDPAPEVPPYTTINVFYEQYLTIPEGL 1140  
Qy 1114 FMLSICLVPTFAVSCLLGLDRLSGLNLLSIVMLVDTVGFMALWDISYNAVSLINLVS 1173  
Db 1141 FMLSICLVPTFAVSCLLGLDRLSGLNLLSIVMLVDTVGFMALWGISYNAVSLINLVS 1200  
Qy 1174 AVGMSVEFVSHITRSFAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGLVILGLAKA 1233  
Db 1201 AVGMSVEFVSHITRSFAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGLVILGLAKA 1260  
Qy 1234 LIQIFFFRLNLLITLLGLLHGLVFLPVILSVGDPVPALALEOKRAEEA VAAVMVASCP 1293  
Db 1261 LIQIFFFRLNLLITLLGLLHGLVFLPVILSVGDPVPALALEOKRAEEA VAAVMVASCP 1320  
Qy 1294 NHPRSVSTADNIYNHSPSGSIKAGATISNLPNNGROF 1332  
Db 1321 NHPRSVSTADNIYNHSPSGSIKAGATISNLPNNGROF 1359

RESULT 9  
AEB93610  
ID AEB93610 standard; protein; 1359 AA.  
XX AC AEB93610;  
XX DT 06-OCT-2005 (first entry)  
XX DE Human NPC1L1 protein.  
KW Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;  
KW cardiant; vasotropic; hyperlipidemia; atherosclerosis; stroke;  
KW arteriosclerosis; ds.

XX OS Homo sapiens.  
XX PN WQ2005069900-A2.  
XX PD 04-AUG-2005.  
XX 14-JAN-2005; 2005WO-US001469.  
XX 16-JAN-2004; 2004US-0537341P.  
XX (MERI ) MERCK & CO INC.  
XX Garcia-Calvo M;  
XX WPI; 2005-564070/57.  
XX N-PSDB; AEB93609.  
XX GENBANK; AF192522.  
XX Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by contacting NPC1L1 with detectably labeled substituted 2-azetidinone glucuronide and a candidate compound and determining if compound binds to human NPC1L1.  
XX Disclosure; SEQ ID NO 44; 215pp; English.  
XX The invention relates to identifying a ligand of NPC1L1. The method involves contacting human NPC1L1 with a detectably labeled substituted 2-azetidinone glucuronide and a candidate compound and determining whether the candidate compound binds to human NPC1L1. In identifying a ligand of NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10 nM or lower. The detectably labeled substituted 2-azetidinone glucuronide is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands are useful for stimulating or blocking the activity of NPC1L1, and for treating conditions caused or mediated by NPC1L1. It is useful for reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart disease, stroke, or arteriosclerosis. The present sequence represents the amino acid sequence of human NPC1L1 as disclosed under GenBank accession number AF192522.  
XX SQ Sequence 1359 AA;

Query Match 99.5%; Score 6872.5; DB 9; Length 1359;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;  
Qy 1 MAEAGLRCWLLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGLMTLSNVCLSN 60  
Db 1 MAEAGLRCWLLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGLMTLSNVCLSN 60  
Qy 61 TPARKITGDHLLILLOKICPRLYTCPNTPQACCSAKQLVSLASLSITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLYTCPNTPQACCSAKQLVSLASLSITKALLTRCPACSDNF 120  
Qy 121 VNLHCHNTCSNQSFLINVT RVAQLGAGQIPAVVAYEAFYOHSPAESYDSCSRVPAA 180  
Db 121 VNLHCHNTCSNQSFLINVT RVAQLGAGQIPAVVAYEAFYOHSPAESYDSCSRVPAA 180  
Qy 181 ATLAVGTCMGYGSALCNQORWLNFGDTGNGLAPLDITFHLLPQGA VSGIGIQLNEG 240  
Db 181 ATLAVGTCMGYGSALCNQORWLNFGDTGNGLAPLDITFHLLPQGA VSGIGIQLNEG 240  
Qy 241 ARCNESOGDDVATCSQDCAASCPAIARPAQALDSTFYLGMPGSLVLIILLCVFAVVTTI 300  
Db 241 ARCNESOGDDVATCSQDCAASCPAIARPAQALDSTFYLGMPGSLVLIILLCVFAVVTTI 300  
Qy 301 LLVGFRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFPQGWGTWVASWPLILVLSV 360  
Db 301 LLVGFRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFPQGWGTWVASWPLILVLSV 360  
Qy 361 IPVVALAAGLVFTTELTTPDVELWSAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSY 420  
Db 361 IPVVALAAGLVFTTELTTPDVELWSAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSY 420

Qy	421	RYDSLILGPKNFSGILDLDDLLLELELELQERLRHLQVMSPEAQRNISLQDICYAPLNP	DNT	480
Db	421	RYDSLILGPKNFSGILDLDDLLLELELELQERLRHLQVMSPEAQRNISLQDICYAPLNP	DNT	480
Qy	481	SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTPKDCGTAL	AL	540
Db	481	SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTPKDCGTAL	AL	540
Qy	541	SCMADYGAPVFPFLAIGGYKGKDYSEAEALIMTFLSNYPAGDPRLAQAQKLWEEAFLE	EM	600
Db	541	SCMADYGAPVFPFLAIGGYKGKDYSEAEALIMTFLSNYPAGDPRLAQAQKLWEEAFLE	EM	600
Qy	601	RAFQRRMAGMFQVTFMAERSLEDEINRTTAEDLP	IFATSYIVIFLYISIALGSSYSWSRV	660
Db	601	RAFQRRMAGMFQVTFMAERSLEDEINRTTAEDLP	IFATSYIVIFLYISIALGSSYSWSRV	660
Qy	661	MVDSKATLGLGGVAVVLGVAAMAGPFSYLGIRSSVLIVQVPPFLVLSVGADNIFIVLE		720
Db	661	MVDSKATLGLGGVAVVLGVAAMAGPFSYLGIRSSVLIVQVPPFLVLSVGADNIFIVLE		720
Qy	721	YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAICFFLGALTPMPAVRTPALTSG	LA	780
Db	721	YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAICFFLGALTPMPAVRTPALTSG	LA	780
Qy	781	ILDFLQMSAFVALLSLDSKQBSASRLDVCCCKVPQELPPPGQEGGLLGLFFQKAYAP	FL	840
Db	781	ILDFLQMSAFVALLSLDSKQBSASRLDVCCCKVPQELPPPGQEGGLLGLFFQKAYAP	FL	840
Qy	841	LHWITRGVULLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA	PA	900
Db	841	LHWITRGVULLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA	PA	900
Qy	901	YFVTTLGYNFSSAGMNAICSSAGCANNFSTQIKQIYATEPPEOSYLAIPASSWVDD	FDW	960
Db	901	YFVTTLGYNFSSAGMNAICSSAGCANNFSTQIKQIYATEPPEOSYLAIPASSWVDD	FDW	960
Qy	961	LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQPHKYL	PWFNLDRP	1020
Db	961	LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQPHKYL	PWFNLDRP	1020
Qy	1021	NIKCPKGGILAAYSTSVNLTSDGQVLT-----ASRFMAYH		1053
Db	1021	NIKCPKGGILAAYSTSVNLTSDGQVLT-----ASRFMAYH		1053
Qy	1054	KPLKNSQDYTEALRAARELAAMITADLRKVPGTDPAPFVFPYTTITNVFVEQYLTIL	PEGL	1113
Db	1081	KPLKNSQDYTEALRAARELAAMITADLRKVPGTDPAPFVFPYTTITNVFVEQYLTIL	PEGL	1140
Qy	1114	FMLSCLVPTFAVSCLLGLDLRSGLLNLSIVMLIVDTVGFMAWMDISYNAVSLINL	VS	1173
Db	1141	FMLSCLVPTFAVSCLLGLDLRSGLLNLSIVMLIVDTVGFMAWMDISYNAVSLINL	VS	1200
Qy	1174	AVGMSVEFVSHITRSPAISTKTWLERAKEATISMGSAVAGVAMTNLPGLIVLGLAKAQ		1233
Db	1201	AVGMSVEFVSHITRSPAISTKTWLERAKEATISMGSAVAGVAMTNLPGLIVLGLAKAQ		1260
Qy	1234	LQIQIFFRNLNLTITLGLLHGLVFLPVLISYVGPDVNPALAEOKRAEBAVAAMVWASCP		1293
Db	1261	LQIQIFFRNLNLTITLGLLHGLVFLPVLISYVGPDVNPALAEOKRAEBAVAAMVWASCP		1320
Qy	1294	NHPSRVSTADNIYVNHSPFSGISKGAGAISNFLPNNGRQF		1332
Db	1321	NHPSRVSTADNIYVNHSPFSGISKGAGAISNFLPNNGRQF		1359
RESULT 10				
ABG22693				
ID	ABG22693 standard; protein; 1344 AA.			
XX				
AC	ABG22693;			
XX				
DT	18-FEB-2002 (first entry)			

XX	DE	Novel human diagnostic protein #22684.
XX	KW	Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.
XX	OS	Homo sapiens.
PN	XX	WO200175067-A2.
PD	XX	11-OCT-2001.
XX	XX	30-MAR-2001; 2001WO-US008631.
PF	XX	31-MAR-2000; 2000US-00540217.
PR	XX	23-AUG-2000; 2000US-00649167.
XX	XX	(HYSE-) HYSEQ INC.
PA	XX	Drmanac RT, Liu C, Tang YT;
PI	XX	WPI; 2001-639362/73.
XX	DR	N-PSDB; AAS86880.
XX	XX	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.
PT	PT	Claim 20; SEQ ID NO 53052; 103pp; English.
XX	PS	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	SQ	Sequence 1344 AA;
		Query Match 94.6%; Score 6536; DB 4; Length 1344;
		Best Local Similarity 96.7%; Pred. No. 0;
		Matches 1271; Conservative 7; Mismatches 21; Indels 16; Gaps 4;
QY	1	MAEAGLRGWLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60
Db	1	MAEAGLRGWLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60
QY	61	TPARKITGDHLTLLOKICPRLYTGNTQACSAKQLVLEASLSITKALLTRCPACSDNF 120
Db	61	TPARKITGDHLTLLOKICPRLYTGNTQACSAKQLVLEASLSITKALLTRCPACSDNF 120
QY	121	VNLHCHNTCSNQSLFINVTRVAQAGQLPAVVAEYAFQHSFAEQSYDSCSRVRVPA 180
Db	121	VNLHCHNTCSNQSLFINVTRVAQAGQLPAVVAEYAFQHSFAEQSYDSCSRVRVPA 180
QY	181	ATLAVGTMCVGVYSALCNAQRLNFPQDGTGNGLAPLDTFHLLPFGQAVGSGIQPLNEGV 240



Db	181	ATLAVTGTCGVYGSALCNARWLNFGQDTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV	240
Qy	241	ARCNESQDDVATSCDCAACSP--AIARQALDSTFYLGQMPGSLVLIILICSVFVV	298
Db	241	ARCNESQDDVATSCDCAACSPAPRPFSTPPSTW--RQMPGSLVLIILICSVFVV	298
Qy	299	TILLVGRVAPARDKSKVDPKGTSLSDKLSFSTHTLLGQFGQNGTGWASWPLTLLVL	358
Db	299	TILLVGRVAPARDKSKVDPKGTSLSDKLSFSTHTLLGQFGQNGTGWASWPLTLLVL	358
Qy	359	SVIPVVALAGLVFTELTDPVELWSAPNSQARSEKAFHQHFGFPFRTNQVILTAENRS	418
Db	359	SVIPVVALAGLVFTELTDPVELWSAPNSQARSEKAFHQHFGFPFRTNQVILTAENRS	418
Qy	419	SYRYSLLGPNKPSGILDLDLLELLELRLHQLWSPQAQRNLSLODICVAPLNPD	478
Db	419	SYRYSLLGPNKPSGILDLDLLELLELRLHQLWSPQAQRNLSLODICVAPLNPD	478
Qy	479	NTSLYDCINSLLOYFQNNRTLLLTANOTLMGOTSQVDMKDHFLYCANAPLTFKDGATL	538
Db	479	NTSLYDCINSLLOYFQNNRTLLLTANOTLMGOTSQVDMKDHFLYCANAPLTFKDGATL	538
Qy	539	ALSCWADYGAVFPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLWEEAFLE	598
Db	539	ALSCWADYGAVFPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLWEEAFLE	598
Qy	599	EMRAPORMAGMFWTFAERSLDEINRTTAEDLPATSYIIVIFLYISALGYSYSSWS	658
Db	599	EMRAPORMAGMFWTFAERSLDEINRTTAEDLPATSYIIVIFLYISALGYSYSSWS	658
Qy	659	RMVDSKATLGGVAVVAVMAAMGPFSLYTRSSLVILQVVPFLVLSVGADNIFIV	718
Db	659	RMVDSKATLGGVAVVAVMAAMGPFSLYTRSSLVILQVVPFLVLSVGADNIFIV	718
Qy	719	LEYQLRPRPGPREVHIGRALGRVAPSMLLCSLSEACFFLGALTMPAVRTFALTSGL	778
Db	719	LEYQLRPRPGPREVHIGRALGRVAPSMLLCSLSEACFFLGALTMPAVRTFALTSGL	778
Qy	779	AVILDFLOMSAFVALLSLDSKQASRLDVCCCKVQOELPPPGQEGILLGFPQKAYAP	838
Db	779	AVILDFLOMSAFVALLSLDSKQASRLDVCCCKVQOELPPPGQEGILLGFPQKAYAP	838
Qy	839	FLHWTITGGV-----LILFLALFGVLSYSMSCHISVGLDQELALPKDSVLLDYFLFLNRYF	894
Db	839	FLHWTITGGVVPSSQLLFLALFGVLSYSMSCHISVGLDQELALPKDSVLLDYFLFLNRYF	898
Qy	895	EVGAPVYFVTTLGYNFSSEAGMNAICSSAGCNFFSTQKIQVATFEPQOSYIALPASSWV	954
Db	899	EVGAPVYFVTTLGYNFSSEAGMNAICSSAGCNFFSTQKIQVATFEPQOSYIALPASSWV	958
Qy	955	DDFIDLWTFSSCCRLYISGPNKDKCPTVNSLNCCLKNCMSITMGSVRPSVEQFHLYLPW	1014
Db	959	DDFIDLWTFSSCCRLYISGPNKDKCPTVNSLNCCLKNCMSITMGSVRPSVEQFHLYLPW	1018
Qy	1015	FUNDARNIKCPKGGLLAAVSTSVNLTSQGOVLASRFWAYHKLKNSQDYTEALRAARELAA	1074
Db	1019	FUNDARNIKCPKGGLLAAVSTSVNLTSQGOVLASRFWAYHKLKNSQDYTEALRAARELAA	1078
Qy	1075	NITADLRKVPGTDPFAPEVPPYITTNVFEQYLTILPEGLFMLSCLVPTFAVSCLLGLD	1134
Db	1079	NITADLRKVPGTDPFAPEVPPYITTNVFEQYLTILPEGLFMLSCLVPTFAVSCLLGLD	1138
Qy	1135	IRSGLLNLLSIYMLVDVTGFWALWDI SYNANVLNLSAVGMSVEFVSHITRSFAISTK	1194
Db	1139	IRSGLLNLLSIYMLVDVTGFWALWDI SYNANVLNLSAVGMSVEFVSHITRSFAISTK	1198
Qy	1195	PWLERAKKATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIFFRNLMLTLLGLLHG	1254
Db	1199	PWLERAKKATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIFFRNLMLTLLGLLHG	1258
Qy	1255	LVFLPVILSVGPDVNNPALALBOKRABEA-----VAAVMVASCNPNHPSRVST	1301
Db	1259	LVFLPVILSVGPDVNNPALALBOKRABEGGSHGGPLAQIITPPESPQITSMST	1313

RESULT 11

ADJ27295

ID ADJ27295 standard; protein; 1331 AA.

XX

ADJ27295;

AC

DT 20-MAY-2004 (first entry)

XX

DE Rat NPC1L1.

XX

KW Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network; plasma membrane; transport signal; promoter;

KW sterol regulated element binding protein 1; SREBP1;

KW binding consensus sequence; transmembrane domain; sterol-sensing domain;

KW SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;

KW cholesterol absorption; serum cholesterol; hyperlipidaemia;

KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.

XX

OS Rattus sp.

XX

PN W02004009772-A2.

XX

PD 29-JAN-2004.

XX

PF 17-JUL-2003; 2003WO-US022467.

XX

PR 19-JUL-2002; 2002US-0397442P.

XX

PA (SCHE ) SCHERING CORP.

XX

PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;

DR WPI; 2004-132945/13.

DR N-PSDB; ADJ27294.

XX

PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting or identifying antagonists of NPC1L1 for inhibiting intestinal cholesterol absorption in a subject, or for treating elevated serum cholesterol or stroke.

PT

XX

PS Claim 1; SEQ ID NO 2; 125pp; English.

XX

CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1 (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains a motif ADJ2731 which acts as a trans-golgi network to plasma membrane transport signal, and which exhibits limited tissue distribution and gastrointestinal abundance. The human NPC1L1 promoter sequence contains a sterol regulated element binding protein 1 (SREBP1) binding consensus sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-sensing domain (SSD) which is involved in sensing cholesterol levels, possibly by a mechanism which involves direct cholesterol bonding. NPC1L1 has 44 amino acid sequence homology to human NPC1, a receptor responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or polynucleotides are useful for detecting or identifying antagonists of NPC1L1, which can be used for inhibiting intestinal cholesterol absorption in a subject, or for treating medical conditions including elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary heart disease, stroke or arteriosclerosis.

CC

XX

SQ Sequence 1331 AA;

Query Match 78.5%; Score 5421.5; DB 8; Length 1331;

Best Local Similarity 77.8%; Pred. No. 0;

Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

Qy 1 MAEAGLRGWLWALLRLIAQSEPTTHIQPGYCAFYECGKNPELSGLTSLNVCSLGN 60

Db 1 MAEAGLRGWLWALLRLIAQSEPTTHIQPGYCAFYECGKNPELSGLTSLNVCSLGN 59

Qy 61 TPARKITGDHLLILQKICPRLYTGNTQ-ACCSAKQLVSLASLSITKALLTRCPACSDN 119

Db 60 TPAHVHTGEHALLQRICPRLYNGPNTTFACCSKTKQLLSBSSMTIKALLTRCPACSDN 119  
QY 120 FVNHLCHNTCSPNQSLFINVTRVAQLGAGQLPAVYAEAFYQHSFAEQSDSCSRVRVPA 179  
Db 120 FVSLHCHNTCSPNQSLFINVTRVVVERGAGEPPAVYAEAFYQSRFAEKAYESCQVRIPA 179  
QY 180 AANTLVGTCMGVYGSALCNAQRWLNFGQDGTGNGLAPLDITFHLLEPQOAVGSGIQLNEG 239  
Db 180 AASLVGSMGVYGSALCNAQRWLNFGQDGTGNGLAPLDITFHLLEPQOALPDGIQPLNGK 239  
QY 240 VARCNESQGDVDDATCSQDCAASCAPATARPQALDSTFYLQMPGSLVLIILCSVPAVVT 299  
Db 240 IAPCNESQGDSDSVCSCQDCAASCVPITPPPEARLPSPFYMGMPQWALIIIFVAVFVLS 299  
QY 300 ILLVGRFVAPARDKSMVDPKGTSLSDKLSFSTHTLLGQFFQCGMTWVASWELTILVLS 359  
Db 300 AVLVLRLVSNRNKNAEGPQAPKLPKHKLSPHTILGRFFQNGWGRTRVASWELTILVLS 359  
QY 360 VIPVVALAAGLVTELTDPVELWSAPNSQARSEKAFHDQHPFPFRTRNQVILITAPNRSS 419  
Db 360 FIVVIALAAGLTFIELTDPVELWSAPKSOARKEKSPHDEHFGFFRTRNQIFVTARNRSS 419  
QY 420 YRYDSLGLGPNFSGILDLLELLELQERLRLHQLVMSPEAQRNLSLODICVAPLNPDN 479  
Db 420 YKYSLLGLGKNFSGILSDFLLELLELQERLRLHQLVMSPEAQRNLSLODICVAPLNPN 479  
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANTQTLGQTSQVDMKDHFLYCANAPLTFKDGITALA 539  
Db 480 TSLSDCCVNSLLQYFQNNRTLLMLTANTQTLGQTSVDMKDHFLYCANAPLTFKDGTSLA 539  
QY 540 LSCMADYGAPVFPFLAIGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLWEEAFLEE 599  
Db 540 LSCMADYGAPVFPFLAVGGYQGYDYSEAEALITFTSLNYPADDDPRMAQAKLWEEAFLEE 599  
QY 600 MRAFORMAGMFOVTTAERSLEDEINRTTAEDLPITFATSYIVIFLYISLALGYSWSMR 659  
Db 600 MESFQNTSKDFQVAFSAERKSEDEINRTTIQDLPFVAFSYIIVFYLISLALGYSRCR 659  
QY 660 VMVDSKATLGLGVAVVLGAVMAAGFFSYLIGRISLVLIIQVVPFLVLSVGADNIFVL 719  
Db 660 VAVESKATLGLGVAVVLGAVLAAGFYSLYGVPSLVLIIQVVPFLVLAAGDNIIFVL 719  
QY 720 EYQRLPRRPGEPREHIGRALGRVAPSMLLCSSEAI CFFLGLALTMPAVRTPALTSGLA 779  
Db 720 EYQRLPRMPGEQREAHIGRTLGSVAPSMLLCSSEAI CFFLGLALTMPAVRTPALTSGLA 779  
QY 780 VILDFLQMSAFVALLSLDSKREASRLDVCCKVQDELPPPGQGLLGFQKAYAPF 839  
Db 780 IILDFLQMTAFVALLSLDSKREASRPDLVCCFSTRKUPPPPKKESGLLRLFRKLIYAPF 839  
QY 840 LLHWITRGVVLLFLALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFVVGAP 899  
Db 840 LLHRFTRPVVMLLFLTLFGANLYLMCNINVLGDQELALPKDSYLLIDYFLFLNRYLEVGP 899  
QY 900 VIFVTTLGNFSEAGNNAICSSAGCNFSTFKIQIYATEFPQSYLAIPASSWDDFID 959  
Db 900 VIFVTTSFGNFSEAGNNAICSSAGCKFSLTQKIYASEFPDQSYVAIAASSWDDFID 959  
QY 960 WLTP -SSCCRLYISGNKDKFCBPSTVNSLNCNKSITMGSVRPSVQFHKYLPFLND 1018  
Db 960 WLTPSSCCRLYIRGHKDEFCBPSTDTSNCLKNCNRTLGPVRPTAEQFHKYLPFLND 1019  
QY 1019 RPNKCPKGLAAYSTSVNLTDQVLAISRFMAYHKPLKNSQDYTEALAAARELANITA 1078  
Db 1020 PPNIRCPKGLAAYRTSVNLSDDQVIAAQFMAHYHPLRNSQDFTALRASRLAANIITA 1079  
QY 1079 DLKRVGTDPAFEPFYTITNVFYEQYLTLPBGLPMLSLCVPTTFAVSCLLGLDLRSG 1138  
Db 1080 DLKRVGTDPEFEPFYTITSNFYQYLTLPBGIPTTALCFVPTFVVCVLLGLDMCSG 1139  
QY 1139 LLNLLSIIMLVDTVCFMADWDISYNAVSLINLVSAGVSEFVSHITSPALSTKPTWL 1198  
Db 1140 ILNLLSIIMLVDTIGLMAVWGISYNAVSLINLVTAVGMSVEFVSHITSPAVSTKPTRL 1199

QY 1199 ERAKEATISMGSAVFAGVAMTNLPGLVLGLAKAQILQIFFFFRNLILTLGLLHGLVFL 1258  
Db 1200 ERAKDATVFMGSAVFAGVAMTNFPFGLILGLFAQAQLIQIFFFFRNLILTLGLLHGLVFL 1259  
QY 1259 FVTLSYVGDVNPALALEQKBAEEAVANVVASCPNHPSRVSTADNIIYNHSEFSGSI-KG 1317  
Db 1260 FVLSYLGPDVNLQALVQEEKLASEA-AVAPEPSCQPYPSPADADAN--VNYGFAPELAHG 1316  
QY 1318 AGAISNLFNNGRQF 1332  
Db 1317 ANAARSLPKSDQKF 1331

RESULT 12  
ADY60738

ID ADY60738 standard; protein; 1331 AA.

XX ADY60738;

XX 19-MAY-2005 (first entry)

XX Rat Niemann-Pick C1-like protein 1 (NPC1L1) protein.

XX Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;  
KW Cardiant; Vasotropic; Cerebroprotective; Cholesterol; hyperlipidemia;  
KW metabolic disorder; atherosclerosis; cardiovascular disease;  
KW coronary artery disease; cerebrovascular ischemia; neurological disease;  
KW arteriosclerosis; transgenic animal.

XX Rattus sp.

XX OS WO2005015988-A1.

XX 24-FEB-2005.

XX 16-DEC-2003; 2003WO-US040113.

XX 17-JUL-2003; 2003US-00621758.

XX 22-AUG-2003; 2003US-00646301.

XX 16-SEP-2003; 2003US-00663208.

XX (SCHE ) SCHERING CORP.

XX Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;

XX WPI: 2005-284403/29.

XX N-PSDB; ADY60737, ADY60746.

PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of  
PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for  
PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.

XX Claim 30; SEQ ID NO 2; 146pp; English.

XX The invention relates to an isolated Niemann-Pick C1-like protein 1  
(NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an  
XX amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,  
XX human and Mouse NPC1L1). Also included are an isolated polynucleotide  
XX encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a  
XX recombinant vector comprising the polynucleotide, a host cell comprising  
XX the vector, an isolated antibody which specifically binds to NPC1L1 (or  
XX to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),  
XX preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous  
XX mutation of endogenous chromosomal NPC1L1 (where the mouse does not  
XX produce any endogenous NPC1L1 protein), an offspring or progeny of the  
XX mouse having inherited a mutated NPC1L1 allele of the mouse, screening a  
XX sample for an intestinal sterol or Salpha-sterol absorption antagonist,  
XX inhibiting NPC1L1 mediated sterol or Salpha-sterol uptake in a subject, a  
XX kit (comprising ezetimibe or substituted azetidinone in a pharmaceutical  
XX dosage form, and information indicating that NPC1L1 is a target of  
XX ezetimibe or substituted azetidinone), decreasing the level of intestinal  
XX sterol or Salpha-sterol absorption in a subject (involving reducing the

CC level of expression of NPC1L1 in the subject), identifying an antagonist  
 CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a  
 CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist  
 CC of NPC1L1 which is useful for inhibiting or decreasing the level of  
 CC NPC1L1 mediated sterol or Salpha-stanol uptake in a subject such as  
 CC mouse, rat or human. The level of expression of NPC1L1 in the subject is  
 CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is  
 CC useful for screening a sample for intestinal sterol or Salpha-stanol  
 CC absorption antagonist. The NPC1L1 antagonists identified are useful for  
 CC treating medical conditions caused or mediated by NPC1L1, e.g.,  
 CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or  
 CC arteriosclerosis. The present sequence represents a mammalian NPC1L1  
 CC protein.

XX SQ Sequence 1331 AA;

Query Match 78.5%; Score 5421.5; DB 9; Length 1331;  
 Best Local Similarity 77.8%; Pred. No. 0;  
 Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGLRGWLWALLRLAQAEPYTHIQPGYCAFYDECGKNPBLSGSLMTLSNVCLSN 60  
 Db 1 MAEAGLRGWLWALLRLAQAEPYTHIQPGYCAFYDECGKNPBLSGSLMTLSNVCLSN 59  
 QY 61 TPARKITGDHLLILKQICPRLYTGPNTQ-ACCSAKQLVSLRASLITKALLTRCPACSDN 119  
 Db 60 TPARKITGDHLLILKQICPRLYTGPNTQ-ACCSAKQLVSLRASLITKALLTRCPACSDN 119  
 QY 120 FVNHLCHNTCPNOSLFINVTRVAGLQAGQPAVVAEAFYQHSFAEQSDYSCSRVRVPA 179  
 Db 120 FVNHLCHNTCPNOSLFINVTRVAGLQAGQPAVVAEAFYQHSFAEQSDYSCSRVRVPA 179  
 QY 180 AATLAVGTCMGYSGALCNAQRWLNFGQDTNGLAFLDITFHLLPQAGVSGIOTPLNEG 239  
 Db 180 AATLAVGTCMGYSGALCNAQRWLNFGQDTNGLAFLDITFHLLPQAGVSGIOTPLNEG 239  
 QY 240 VARCNESQGDVATCSCODCAACPAIARPOLADSTFVLGMPGSLVLIILICSVFAVVT 299  
 Db 240 VARCNESQGDVATCSCODCAACPAIARPOLADSTFVLGMPGSLVLIILICSVFAVVT 299  
 QY 300 ILLVGRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVASWPLTILVLS 359  
 Db 300 ILLVGRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVASWPLTILVLS 359  
 QY 360 VIPVVALAAGLVFTLTDTPVELWSAPNSQARSEKAFHQDHGPPFRFNQVILTAPESS 419  
 Db 360 VIPVVALAAGLVFTLTDTPVELWSAPNSQARSEKAFHQDHGPPFRFNQVILTAPESS 419  
 QY 420 YRYDSLILGPKNFSGILDLDLLELLELQERLHLOWNSPEAQRNLSQDICYAPLNPDN 479  
 Db 420 YRYDSLILGPKNFSGILDLDLLELLELQERLHLOWNSPEAQRNLSQDICYAPLNPDN 479  
 QY 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGTLA 539  
 Db 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGTLA 539  
 QY 540 LSCMADYCAPPPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEAEFLKE 599  
 Db 540 LSCMADYCAPPPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEAEFLKE 599  
 QY 600 MRAFORMMAGFQVTFTAERSLEDEINRTTAEPLFIPATSYIVIFLYISLALGYSWSR 659  
 Db 600 MRAFORMMAGFQVTFTAERSLEDEINRTTAEPLFIPATSYIVIFLYISLALGYSWSR 659  
 QY 660 VMVDSKATLGLGGVAVLGVAMAMGFFSYLGRSSVLVQVVPFLVLSVGADNIFIVL 719  
 Db 660 VMVDSKATLGLGGVAVLGVAMAMGFFSYLGRSSVLVQVVPFLVLSVGADNIFIVL 719  
 QY 720 EYQRLPRRGPGEVHIGRALGRVAPSMLLCSLSEACFFLIGALTPMPAVRTFALTSLGA 779  
 Db 720 EYQRLPRRGPGEVHIGRALGRVAPSMLLCSLSEACFFLIGALTPMPAVRTFALTSLGA 779  
 QY 780 VILDFLLQMSAFVALLSLDSKQREASRLDVCCVKPQBLPPPGQEGLLGFFQKAYAPF 839

Db 780 IILDFLLQMSAFVALLSLDSKQREASRPDLVLCFSTFKLPPPKKEGULLRFFRKIYAPF 839  
 QY 840 LLHWITRGVLLFLALFGVSLYSMSCHISVGLDQELALPKDSYLLDYFLNRYFEVGP 899  
 Db 840 LLHWITRGVLLFLALFGVSLYSMSCHISVGLDQELALPKDSYLLDYFLNRYFEVGP 899  
 QY 900 VYFVTTLGYNFSSEAGNNAICSSAGCNFNFTQKIQYATFEPQSYLAIASSWVDDFID 959  
 Db 900 VYFVTTLGYNFSSEAGNNAICSSAGCNFNFTQKIQYATFEPQSYLAIASSWVDDFID 959  
 QY 960 WLTP-SSCCRLYISGPNKDFCPSVNSLCLNCLNCSITMGSVRPSVEQFHKYLWFLND 1018  
 Db 960 WLTP-SSCCRLYISGPNKDFCPSVNSLCLNCLNCSITMGSVRPSVEQFHKYLWFLND 1018  
 QY 1019 RPNIKCPKGLAAYSTSVNLTSDQVLAASFMAHFKPLKNSQDYTEALRAARELANITA 1078  
 Db 1020 RPNIKCPKGLAAYSTSVNLTSDQVLAASFMAHFKPLKNSQDYTEALRAARELANITA 1078  
 QY 1079 DLKRVGCTDPAFEVFPYITNVFVEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLRS 1138  
 Db 1080 DLKRVGCTDPAFEVFPYITNVFVEQYLTILPEGLFMLSCLVPTFAVSCILLGLDLRS 1138  
 QY 1139 LMLLSIVMLIVDTVGFMAWLDISYNAVSLINLVSAGMSVFEVSHITRSFAISKPTWL 1198  
 Db 1140 LMLLSIVMLIVDTVGFMAWLDISYNAVSLINLVSAGMSVFEVSHITRSFAISKPTWL 1198  
 QY 1199 ERAKEATISMGSAVAVGAVMTNLPGLILVLGAKAQLIQIFFRNLNLLITLLGLHLVFL 1258  
 Db 1200 ERAKEATISMGSAVAVGAVMTNLPGLILVLGAKAQLIQIFFRNLNLLITLLGLHLVFL 1258  
 QY 1259 PVILSVYGVDPVNPALALEQKRAEAAVAVASCPNHPRSRVSTADNLYVNHSPGSI-KG 1317  
 Db 1260 PVILSVYGVDPVNPALALEQKRAEAAVAVASCPNHPRSRVSTADNLYVNHSPGSI-KG 1317  
 QY 1318 AGAISNLPNNRQF 1332  
 Db 1317 ANAARSLPKSDQKF 1331

RESULT 13  
 AEB93568  
 ID AEB93568 standard; protein; 1331 AA.  
 XX AEB93568;  
 AC AEB93568;  
 XX 06-OCT-2005 (first entry)  
 XX Rat NPC1L1 protein.  
 XX Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;  
 XX cardiant; vasotropic; hyperlipidemia; atherosclerosis; stroke;  
 XX arteriosclerosis; rat.  
 XX Rattus sp.  
 XX WO2005069900-A2.  
 XX PD 04-AUG-2005.  
 XX 14-JAN-2005; 2005WO-US001469.  
 XX 16-JAN-2004; 2004US-0537341P.  
 XX (MERI ) MERCK & CO INC.  
 XX Garcia-Calvo M;  
 XX WPI; 2005-564070/57.  
 XX N-PSDB; AEB93567.  
 XX Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by  
 XX contacting NPC1L1 with detectably labeled substituted 2-azetidinone

PT glucuronide and a candidate compound and determining if compound binds to  
XX human NPC1L1.

Example 2; SEQ ID NO 2; 215pp; English.

CC The invention relates to identifying a ligand of NPC1L1. The method  
CC involves contacting human NPC1L1 with a detectably labeled substituted 2-  
CC azetidinone glucuronide and a candidate compound and determining whether  
CC the candidate compound binds to human NPC1L1. In identifying a ligand of  
CC NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10  
CC nM or lower. The detectably labeled substituted 2-azetidinone glucuronide  
CC is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands  
CC are useful for stimulating or blocking the activity of NPC1L1, and for  
CC treating conditions caused or mediated by NPC1L1. It is useful for  
CC reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart  
CC disease, stroke, or arteriosclerosis. The present sequence represents a  
CC rat NPC1L1, a N-glycosylated protein.

XX Sequence 1331 AA;

Query Match 78.5%; Score 5421.5; DB 9; Length 1331;  
Best Local Similarity 77.8%; Pred. No. 0;  
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY	1	M A P A G L R G W L L W A L L L R L A Q S E P Y T T H O P G Y C A F V D E C K N P E L S G L S M T L S N V S C L S N 60
DB	1	M A A A W L - G M L L W A L L S A A Q G E L Y T P K H E A G V C T F Y E E C G K N P E L S G L S T S L S N V S C L S N 59
QY	61	T P A R K I T G H L L I L L O K I C P R L Y T G P N T Q - A C C S A K O L V S L E A S L S I T K A L L T R C P A C S D N 119
DB	60	T P A R H V T G E H L A L L Q R I C P R L Y N G P N T T A C C S T K Q L L S L E S S M S I T K A L L T R C P A C S D N 119
QY	120	F V N L H C H N T C S P N Q S L F I N V R V A O L G A G O L P A V V A Y E A F Y Q S H P A E Q S V D S C S R V R V P A 179
DB	120	F V S L H C H N T C S P Q S L F I N V R V V E R G A G E P P A V V A Y E A F Y Q S P A E K A Y E S C S Q V R I P A 179
QY	180	A A T L A V G T W C G Y G S A L C N A Q R W L N F Q G D T G N G L A P L D I T F H L L E P Q A V G S G I O P L N G S 239
DB	180	A A S L A V G S C G Y G S A L C N A Q R W L N F Q G D T G N G L A P L D I T F H L L E P Q A L P D G I O P L N G K 239
QY	240	V A R C N E S Q G D V A T C S Q C A C S P A I A R P O A L D S T F Y L G O M P G S L V L I I L C S V E A V T 299
DB	240	I A P C N E S Q G D S A V C S Q C A C S P V I P P E A L R P S F Y M G R M P G W L A I I I T F A V E V L L S 299
QY	300	I L L V G R V A P A R D K S R W D P K K T S L S D K L S F S T H L L G O F Q C G W T V A S W P L T L I V L S 359
DB	300	A V L V R L R V S N R N K N A E G P Q E A P K L P H K H L S P H T I L G R F F Q N G W T R V A S W P L T V L A S 359
QY	360	V I P V A L A A G L V P T E L T T D P V E L W S A P N S Q A R S E K A F H D O H G P F F R T N Q V I L T A P N R S S 419
DB	360	F I V V I A A A G L T F I E L T T D P V E L W S A P K S Q A R K E S F H D E H G P F F R T N Q I F V T A R N R S S 419
QY	420	Y R Y D S L L G P K N F S G I L D L L L E L E L O E R L H R L Q V W S P E A Q R N I S L O D I C Y A P L N P D N 479
DB	420	Y K Y D S L L L G S K N F S G L S D F L L E L E L Q E R L H R L Q V W S P E A R N I S L Q D I C Y A P L N P Y N 479
QY	480	T S L Y D C C I N S L L Q Y F O N N R T L L L T A N Q T L M G T S O V D W K D H F L Y C A N A P L T F K D G T A L A 539
DB	480	T S L S D C C V N S L L Q Y F O N N R T L L M L T A N Q T I N G T S L V D W K D H F L Y C A N A P L T F K D G T S L A 539
QY	540	L S C M A D Y G A P V P F F L A I G Y K G K O Y S E A B A L I M T F S L N N Y P A G D P R L A Q A K L W E E A F L E E 599
DB	540	L S C M A D Y G A P V P F F L A V G Y Q G T D Y S E A B A L I I T F S L N N Y P A D D P R M A Q A K L W E E A F L K E 599
QY	600	M E A F O R M A G M Q V T T A E R S L E D E I N R T A E D L P I F A T S Y I V I F L Y I S L A L G S Y S S W S R 659
DB	600	M E S F Q R N T S D K F Q V A F S A E R S L E D E I N R T T I Q D L P V F A V S Y I I V F L Y I S L A L G S Y S R C S R 659
QY	660	V M V D S K A T I L G L G V A V V L G A V M A M G F F S Y L G I R S S L V I Q V V P F L V L S V G A D N I F I V L 719
DB	660	V A V E S K A T I L G L G V I V L G A V L A M G F Y S Y L G V P S S L V I I Q V V P F L V L A V G A D N I F I V L 719
QY	720	E Y Q R L P R P G E P R E V H I G R A L G R V A P S M L L S E A I C F F L G A L T P M P A V R T F A L T S G L A 779

DB	720	E Y Q R L P R M P G E O R E A H I G R T L A S V A P S M L L S E A I C F F L G A L T P M P A V R T F A L T S G L A 779
QY	780	V I L D F L L O M S A F V A L L S L D S K R Q E A S R L D V C C V K P Q E L P P P Q G G L L L G P F Q K A Y A P F 839
DB	780	I I L D F L L Q M T A F A L L S L D S K R Q E A S R P D V L C C F S T R K L P P P P K E G L L L R F F R K I Y A P F 839
QY	840	L L H W I T R G V V L L L F L A L F G V S L Y S M C H I S V G L D Q E L A L P K D S Y L L D Y F L F L N R Y F E V G A P 899
DB	840	L L H R F I R P V V M L L F L T F G A N L Y L M C N I N V G L D Q E L A L P K D S Y L I D Y F L F L N R Y L E V G P P 899
QY	900	V F V T T L G N F S E A G M N A I C S S A G C N F S F T Q I O Y A T E F P E Q S Y L A I P A S S W V D D F I D 959
DB	900	V Y F V T T S G F N F S E A G M N A T C S A G C K S F S L T Q I O Y A S E F P D Q S Y V A I A A S S W V D D F I D 959
QY	960	W L T P - S S C R L Y I S G N K D K F C P S T V N S L N C L K N C M S I M T G S V R P S V E O P H K Y L P W F L N D 1018
DB	960	W L T P S S C R L Y I R G P H K D E F C P S T D T S F N C L K N C M R T L G P V R P T A E Q F H K Y L P W F L N D 1019
QY	1019	R P N I R C P K G L A A Y S T S V N L T S D G V L A S R F M A Y H K P L K N S O D Y T E A L R A A R E L A A N I T A 1078
DB	1020	P E N I R C P K G L A A Y R T S V N L S D G Q V I A S Q F M A Y H K P L A N S Q D F T E A L R A S R L L A A N I T A 1079
QY	1079	D L R K V P G T D P A E V E P Y T T I T N V F Y E Q Y L T I L P E G L P M L S C L V P T F A V S C L L L G L D L R S G 1138
DB	1080	D L R K V P G T D P N F E V P Y T T I S N V F Y Q O Y L T V L P E G I F T L A L C F V P T F V W C V L L L G L D W C S G 1139
QY	1139	L N L L S I V M L V D T V G F M A L W D I S Y N A V S L I N L S A V G M S V F S H I T S E A I S T K P T W L 1198
DB	1140	I L N L S I M I L V D T I G L M A V M G I S Y N A V S L I N L V A V G M S V F S H I T S E A V S T K P T R L 1199
QY	1199	E R A K E A T I S M G S A V A G A M T N L P G I L V L G L A K A Q I O I F F R L N L L I T L L G L L H G L V E L 1258
DB	1200	E R A K D A T V M G S A V A G A M T N F P G I L L G F A Q A Q I O I F F R L N L L I T L L G L L H G L V E L 1259
QY	1259	P V I L S V G P D V N P A L A E O K R A E A V A A V M V A S C N P H S R V S T A D N I Y N H S P E S G I - X G 1317
DB	1260	P V V L S Y L G P D V N Q A L V Q E E K L A S E A - A V A P E S C P Q Y P S P A D A D A N - - V N Y G F A P E L A H G 1316
QY	1318	A G A I S N F L P N N G R O F 1332
DB	1317	A N A R S S U P K S D Q K F 1331
RESULT 14		
ID	ADJ27305	ADJ27305 standard; protein; 1333 AA.
XX	AC	ADJ27305;
XX	DT	20-MAY-2004 (first entry)
XX	DE	Mouse NPC1L1.
XX	KW	Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;
KW	KW	plasma membrane; transport signal; promoter;
KW	KW	sterol regulated element binding protein 1; SREBP1;
KW	KW	binding consensus sequence; transmembrane domain; sterol-sensing domain;
KW	KW	SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;
KW	KW	cholesterol absorption; serum cholesterol; hyperlipidemia;
XX	XX	atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
OS	Mus sp.	
XX	PN	WO2004009772-A2.
XX	XX	
XX	XX	29-JAN-2004.
PD	PD	
PF	PF	17-JUL-2003; 2003WO-US022467.
XX	XX	
PR	PR	19-JUL-2002; 2002US-0397442P.
XX	XX	
PA	PA	(SCHE ) SCHERING CORP.





Qy	1262	LSYVGPDVNEPALAEOKRAEEAAVAVMVASCPNHPSRVSTADNIIYVNHSEGS- IKGAGA	1320
Db	1263	LSYLGPDVNOALVLEEKATEA-AMVSEPCQIYPPADANTSDYVNYGENPEFIPEINA	1321
Qy	1321	ISNFLPNNGRQF	1332
Db	1322	ASSSLPKSDQKF	1333

Search completed: April 7, 2006, 19:16:33  
Job time : 208 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:16:49 ; Search time 52 Seconds  
(without alignments)  
2464.629 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 M8AGLGRGWLWALLRLAQ.....GSIKGAGATSNFLPNNGRQF 1332

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2385	34.5	1278	2 T30188	Niemann-Pick C dis
2	1394	20.2	1055	2 T05663	hypothetical prote
3	1347	19.5	1456	2 T15961	hypothetical prote
4	1329	19.2	1170	2 S53525	probable membrane
5	646.5	9.4	1434	2 T30172	transmembrane prot
6	628.5	9.1	1442	2 T18538	patched protein -
7	600	8.7	915	2 S44797	F09G8.4 protein -
8	598.5	8.7	933	2 T25600	hypothetical prote
9	583.5	8.4	1182	2 T13952	membrane protein p
10	573.5	8.3	1003	2 T26746	hypothetical prote
11	572	8.3	1220	2 T18291	patched protein -
12	539.5	7.8	889	2 T29590	hypothetical prote
13	514	7.4	1405	2 T27969	hypothetical prote
14	473.5	6.9	1015	2 T15830	hypothetical prote
15	464.5	6.7	1299	1 S06119	membrane protein p
16	458.5	6.6	955	2 T21612	hypothetical prote
17	436	6.3	956	2 A89153	protein C24B5.3 (i
18	409.5	5.9	936	2 T26521	hypothetical prote
19	407	5.9	800	2 T26683	hypothetical prote
20	399.5	5.8	881	2 T31739	hypothetical prote
21	381	5.5	840	2 T33217	hypothetical prote
22	379	5.5	714	2 T16126	hypothetical prote
23	367	5.3	690	2 T23399	hypothetical prote
24	367	5.3	877	2 T24097	hypothetical prote
25	349	5.1	845	2 T25657	hypothetical prote
26	347.5	5.0	820	2 T32908	hypothetical prote
27	312	4.5	890	2 T22186	hypothetical prote
28	297.5	4.3	871	2 T28706	hypothetical prote
29	294	4.3	983	2 T21213	hypothetical prote

30	276.5	4.0	633	2 S44795	F09G8.3 protein -
31	259.5	3.8	413	2 S28276	hypothetical prote
32	224	3.2	1276	2 T18526	SREBP cleavage act
33	189.5	2.7	1154	2 T48829	related to SREBP c
34	187.5	2.7	932	2 T28820	hypothetical prote
35	186.5	2.7	1227	2 T20370	hypothetical prote
36	179.5	2.6	787	2 H71453	hypothetical prote
37	173	2.5	969	2 T33156	hypothetical prote
38	170	2.5	746	2 A75018	transport protein
39	154	2.2	823	2 B81282	probable ABC-type
40	149	2.2	1011	2 T07712	probable ABC-type
41	149	2.2	1051	2 AG3455	multidrug resistan
42	148.5	2.1	746	2 H84301	hypothetical prote
43	146	2.1	1041	2 AC0423	multidrug efflux p
44	143	2.1	724	2 H69780	antibiotic transpo
45	141.5	2.0	1049	2 AF0561	acriflavin resista

ALIGNMENTS

RESULT 1

T30188  
Niemann-Pick C disease protein - mouse  
N;Alternate names: NCPI protein  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30188  
R;Loftus, S.K.; Morris, J.A.; Carstea, E.D.; Gu, J.Z.; Cummings, C.; Brown, A.; Ellis-Science 277, 232-235, 1997  
A;Title: Murine model of niemann-pick C disease: mutation in a cholesterol homeostasis A;Reference number: Z20765; MUID:97362324; PMID:9211850  
A;Accession: T30188  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1278 <LOF>  
A;Cross-references: UNIPROT:O35604; UNIPARC:UPI00000299BC; EMBL:AF003348; NID:G2251241 C;Genetic:  
A;Gene: Npc1  
A;Map position: 18

Query Match 34.5%; Score 2385; DB 2; Length 1278;  
Best Local Similarity 40.2%; Pred. No. 1.2e-158;  
Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;

QY	14	LLRLAQSEPYTTIHQPGYCAFVDECGKNPELGGSLMTLSNVSLNTSPARKITGDHLIL	73
DB	12	LLLLCPAQVFSQ-----SCVWIGCG---IATGD---KRYNCKYGGPKPLPKDGYDL	59
QY	74	LQKICPRLYTGNTQACCSAKQLVSLASLSITKALLTRCPACSDNFVNLHCHNTCSPNQ	133
DB	60	VQELCPGLFF-DNVSICCIDIQLQTLKSNLQLPLQFLSRCPSCFYNNLTLFCELTCSPHQ	118
QY	134	SLFINVTRVAQIAGQGLPA-----VVAYEAFYQHSFAEQSYDCSRRVVRPAATLAVGTWC	189
DB	119	SOFLNVATATEDFDPKTPENKTNVKELEYVQGSFANAMYNACRDVEAFSSNEKALGLLC	178
QY	190	GVYGSALCNAQWLNPFQDGTGNGLAPLDI-----TFHLEPGQAVSGSQPLNEGVARCN	244
DB	179	GDARA-CNATWIEYMFNKGNGQAPFTIIPVFSLSIL-----GMEPRNRNATKGCN	229
QY	245	ESQGDVATPCQDCAASC-----PAIARPOALDSTF-----YLQMPGSLV	286
DB	230	ESVDEVTGSCQDCSIVCGPKQPQPPPPMPWRIWGLDAMYIMVYVAFLFVFGALL	289
QY	287	LIILCSFVAVVTILLVGRVAPARDKSMVDPKKGTSLSDKLSFSTHTLLLGQFPQGWGT	346
DB	290	AVWCHRRRYFVSEYTPIDSNIAFSVNSS-----DKGEASCCDPLGAFAFDCLARMFTKWGA	345
QY	347	WTASWPLTLLVSLVPVVALAAGLVFTELTTDPVELWSAPNSQARSEKAFHQHFGPFR	406
DB	346	FCVRNPTCLIFPSLAFITVCSGSLVFVQVTINPVELWSAPHSQARLEKEYFKDHFGPFR	405

```
Qy 407 TNQVLTAPNRSSRYRYSLLGPK-NFSGIILDLDLLELLELQERLRHLQVMSPEAQRN- 464
Db 406 TEQLIQAPNTSVHYEPYAGADVFPGLPKELHQLVNLQ-----IATESITASYNN 460
Qy 465 --ISLQDICYAPLPDNTSLDYDCINSILQYFQNNRTLLLTANQTLWGQTSQV----- 516
Db 461 ETVTLQDICVAPLSPYK--NCTIMSVLNTFQNSHAVL-----DSQGDFFYI 506
Qy 517 --DWKDEFLYCANAPLTFKDGTALALSCMADYGAPVFPFLAIGGYKGYDEABALIMTF 574
Db 507 YADVHTHELYCVRAPASLNDTSLHGLCLGTFGGPFWPLVGLGYDDQYNNATALVTF 566
Qy 575 SLNNYPAGDPRLAQAKLWEEAFLEMRAPQRMAGMFQVTTAERSLEDEINRTTAEDLP 634
Db 567 PVNNYNDTERLQRAWAKEFISFVKYKN---PNLTISTAERSIEDELNRESNDSVF 623
Qy 635 IPATSYVIFLYISIALGSYSMSRWVDSKATGLGGVAVVLGAVMAAMGFESVLGIRS 694
Db 624 TVIISVVMFLYISIALGHIOCSRLLVDKISLGILVILSSVACSLGIFSTMGPL 683
Qy 695 SILVILQVVPFLVLSVGADNIFIVLEYQRLPRRPGEPREHVHIGRALGRVAPSMILCSLSE 754
Db 684 TLIVIEVLPFLVAVGVNIFILVQYQDERLQBEETDQQLGRILGEVAPTMFLSSFSE 743
Qy 755 AICFFLGLATPMPAVRTFALTSLGLAVILDFLQMSAFVALLSLDSKROEASRLDVCCCVK 814
Db 744 TSAFFFGALSSMPAVHTFSLFAGMAVLIDFLQITCFVSLGLDITKROEKHLDILCCVR 803
Qy 815 PQELPPPGQ---EGLLGPFQKAYAPFLHWHITRGVVVLLFLALFGVLSYMSCHISVG 870
Db 804 GAD---DQGGHASESYLFRFKNYFAPLLKDWLRPIVAVFVGLVSFSAVNKVVDIG 860
Qy 871 LDQELALPKDSYLLDYFLFLNRYFEVGPVVFVTTLTGYNFSSEAGMNAICSSAGCNSFSF 930
Db 861 LDQSLSPNDSVIANFKSLAQYLHSGPPVYFVLEEGYSSRKQNNVWCGMGCDNDSL 920
Qy 931 TQKIYATEFPQSYALIPASWDDFDWLTP--SSCRLYISGNKDKFCESTVNSLNC 989
Db 921 VQOIENAAELDTYTRVGAPSSWIDDYDFWVSPQSSCCRLY---NVTHQFCNASYMDPTC 977
Qy 990 LKNCMSIT--MGSVRPSQGFHKLPMFLNDRNPKPKGLAAVSTSNLSDGQVL--AS 1047
Db 978 VR-CRPLPEGKQRFQGGKFMFLPMFLNDRNPKPKGLAAVSTSNLSDGQVL--AS 1036
Qy 1048 RFMAYHKPLKNSQDYTEALRAARELAANITADLRKVPDTPAFEPVFPYTIINNVFYEQLT 1107
Db 1037 YFWTYHTILKTSADYTDAMKKARLIASNTITMRS--KGSD--YRVFPYSVFVFEQVLT 1093
Qy 1108 ILPEGLFMLSCLVPTFAVSCLLGLDLSGLNLLSIVMLIVDTVGPWALWDISYNAVS 1167
Db 1094 IIDDITFNLVSLGSIFLVTLVVLGCELWSAVIMCITIAMILVNNMFGVMWMLGISLNAVS 1153
Qy 1168 LNLVLSAVCMSVEFVSHITRSFAISTKPTWLRAXEATISMGSAVAGVAMTNLPGLVL 1227
Db 1154 LNLVNSCGISVEFCSHITRATMTKSGRVSAREALAHMGSSVFSGITLTKFGIVVL 1213
Qy 1228 GLAKAQLIQIFFRNLNLTLLGLHLGVFLFVLVSYYGPDVNP 1272
Db 1214 AFAKQIPEIFYRMYLAWVLGATHGLIFLFPVLLSYIGSPVNYKA 1258
```

## RESULT 2

T05663

hypothetical protein F22I13.120 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C&gt;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004

C:Accession: T05663

R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.

submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15420

A:Accession: T05663

A:Molecule type: DNA

A:Residues: 1-1055 &lt;BV&gt;

A;Cross-references: UNIPROT:Q9SVF0; UNIPARC:UPI000000A52EC; EMBL:AL035539  
A;Experimental source: cultivar Columbia; BAC clone F22I13  
C;Genetics:  
A;Map position: 4  
A;Introns: 24/3; 60/3; 99/3; 150/1; 193/1; 216/3; 278/3; 364/3; 396/3; 41  
026/3  
A;Note: F22I13.120

Query Match 20.2%; Score 1394; DB 2; Length 1055;  
Best Local Similarity 28.5%; Pred. No. 2.6e-89;  
Matches 377; Conservative 221; Mismatches 385; Indels 340; Gaps 41;

```
Qy 36 YDECGKNPELSGSLMTLSNVCSLNTPARKITGDHLI--LLOKTCPLRYTGPNTQAACSA 93
Db 2 YDICGHRSD--GKVL-----NCPYASPS--IQDELFSAKIQSLCPTI-----SGNVCCTE 48
Qy 94 KQLVLSLEASISITKALITRCPAGCDNFVNLHCHINTCSPNQSLFINVTRVLAQLGAGOLPAV 153
Db 49 TQPDTLRSQQVQAVPFLVGGCPACLRNLFNLFCBLSLCSQSLFINVTSVAEV--SGNL-TV 106
Qy 154 VAYEAFYQHSFASQSDSCSRVRVPAATAVGTMCVYGSALCNAORWLNFGQD---TG 210
Db 107 DGIDYHITDTFGEGLYESCKEVKFGTMTNRAINPV-----GGGAKNFREWFTFIGQKAPSG 162
Qy 211 NGLAPLDTIHLLEPGQAVSGIQPLNEGVARCNESQDDVATCSCODCAASCFAIARPQ 270
Db 163 FPGSPYAINPKSSIP-----ESSAMVPMN-----VSVYSCACSSPEPLPPH 203
Qy 271 ALDS--TFYLG-----QMPGSLVILIIILCSFVAVVTILLVGRFVAPARDKSNMVDPKKG 322
Db 204 DEUDCSIRIGPLKVCRIEELSNALVYVLLVSCFCWAGLNRRRNTTQPLDSSKPLHPVEE 263
Qy 323 TSLSDKLSFSHTLLG-----QFFQGWGTWVASWPLTILVLVSVIPVVA 365
Db 264 DGINSEMK--ENILGVKQVORHAQLSPVORMAFYRSYGSWIARNPSLVLFMSVAIVLA 320
Qy 366 LAAGLVTELTDPVELWSAPNSOARSEKAFHDHGFEPFRNTNOVIL--TAPNRSYRYS 424
Db 321 LCSGLYNFKVETREKULWGPESKAEEKFPFDTHLSPFYRIEQLIILATVPDPKSGRAPS 380
Qy 425 LLLGPKNFGSGLDLDLLELELERLHLQVMSPEAQRNLSQDICYAPLPDNTSLYD 484
Db 381 I-----VTDENILL-LFDIQK----- 396
Qy 485 CCINSLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLCYCANAPLTFKDTALALSMA 544
Db 397 -----YFKWD-----SGTFDDYGGVEHAEC-----FQHYTS-SETCLS 429
Qy 545 DYGAVPFPFLAIGGYKGYKDYSE-----AEALIMTFSLNNYP 580
Db 430 AFQAPVDPSPAVLGGFSGNNYSEVMVSELGCSVPFDCYDVKRTLFOATAFVVTYPVNNV- 488
Qy 581 AGDP--RLAQAKLWEEAF-----BEMRAFQRMAGMFQVTTAERSLEDEINRTTAEDLP 634
Db 489 IGDSSNENARAVAMEKSFQIAKEELLPMVR--SKNLSLSPSSSESIIEELKRESTADVI 546
Qy 635 IPATSYVIFLYISIALGSYSMSRWVDSKATGLGGVAVVLGAVMAAMGFESVLGIRS 694
Db 547 TIAASYLVMFVYISVTLGDAPQFYTFYISSKVLGLSGVLLVLLSVLGSVGSALGVKS 606
Qy 695 SILVILQVVPFLVLSVGADNIFIVLEYQRLPRRPGEPREHVHIGRALGRVAPSMILCSLSE 754
Db 607 TLIIMEVIFPLVAVGVNDNMCILVHVKRQPREVS--LEQRISSALVEGFSITLASLSE 664
Qy 755 AICFFLGLATPMPAVRTFALTSLGLAVILDFLQMSAFVALLSLDSKROEASRLDVCCCVK 814
Db 665 VLAFAVGAFVMPACRIFSMFAALIMLDFLQITAFVALIVFDCKRSADNRIDCFPCIK 724
Qy 815 -----PQELPPPGQSEGILLGPFQKAYAPFLHWHITRGVVVLLFLALFGVLSYMSCHISVG 870
Db 725 VPSSRSRESVEGEBFGFLERYMKVEHAPVLGLMGVMVWVAVFFA----- 769
Qy 871 LDQELALPKDSYLLDYFLFLNRYFEVGPVVFVTTLTGYNFSSEAGMNAICSSAGCNSFSF 930
```

```
Db 770 -----:-----FAL 772
Qy 931 TQKIQVATEPPEQSYLAIPASSWVDDFDLWLTSS--CCRLYISGPNKDKFCBPSTVNSLN 988
Db 773 ASIIISASQASDTSYIAKPAASWLDLFLVWLSPEAFGCCCKFTNG-----SYCPP---DDQ 825
Qy 989 CLKNCMSITMGSRVPSVEQFHKYLPWFLNDRPNIKCPKGLAAYSTSVNLT--SDQVLA 1046
Db 826 CFRHS--DLVQDRPSTAQFREKLWFLNALPSADCAKGGHGYTNSVDLKGYESVIOA 882
Qy 1047 SRFMAYHKLKNSQDYTEALRAARELANITADLRKVPGTDPAFEVFPYTIINVFYEQYL 1106
Db 883 SBFRYHTEL-NTQ-----IDIFFYSVFVFPEQYL 912
Qy 1107 TILPEGLFMLSCLVPTFAVSCLLGLDLRLSGLLNLLSTVMILVDTVGFPMALWDISYNVAV 1166
Db 913 NIWTVALTNLAIV-----GIQLNAV 934
Qy 1167 SLINLVSAGMGVEFVSHITRSFAISTKPTMLERAKEATISMGSAVFAGVAMTNLPGILV 1226
Db 935 SVVNLIMSIGIAVEFCVHISHAFLMSSGDR-EHRAREALFTWGA SVFGIITLTLVGIV 993
Qy 1227 LGLAKAQLIQIFFRLNLLITLGLLHGLVFLPVLVSYG-PDVNPALALEOKRAEAVA 1285
Db 994 LCFARSEIFVYFQMYALAVIIGFLHGLVFLPVLISLAGPPQLN--LDIEQQQDTDEASS 1051
Qy 1286 AVN 1288
Db 1052 SLL 1054

RESULT 3
Ti5961
hypotheical protein F02E8.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: Ti5961
R:Miller, N.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F02E8.
A:Reference number: Z18436
A:Accession: Ti5961
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1456 <MIL>
A:Cross-references: UNIPROT:Q19127; UNIPARC:UPI000017B8AD; EMBL:U53340; NID:g1255859; PI
A:Experimental source: strain Bristol N2; clone F02E8
C:Genetics:
A:Gene: CESP:F02E8.6
A:Map position: X
A:Introns: 71/1; 124/3; 218/3; 269/3; 310/3; 403/2; 484/3; 528/2; 764/1; 858/3; 888/3; 1

Query Match 19.5%; Score 1347; DB 2; Length 1456;
Best Local Similarity 27.5%; Pred. No. 7.7e-86;
Matches 372; Conservative 258; Mismatches 565; Indels 158; Gaps 41;

Qy 76 KICRLYTGPTQACCSAKQLVLSIASISITKALLTRCPACSDNFVNLHCNTCSNQSL 135
Db 63 EFCPHLLTGDN-KLCCTPSQAEGLTKQIAQARHILGRCPSCFDNFAKLWCEFTCSNQD 121
Qy 136 PINTRVAQL--GAGQLPVAVAYEAF-----YQHS--FAEQSYDSCSRVRVPAATLAVG 186
Db 122 FVSISEMKPIEKGEGTPEYQPAAYVNTVYRISTDPAGNFSCKDKDVTGGGPALRV- 180
Qy 187 TMCVGYGALCAQARWLNFQDGTGNGL-APLIDITFHLLEPQOAVGSGIQP-LNBEVARCN 244
Db 181 -MC---TSTPCTLTNLWLEFICTQNLDLNIPHTKFLLYDPIKTPPSDRSTVMVNFTEGD 236
Qy 245 ESQGDVATCSQDCAACPAIRAPQALDSTFYLGQ-----PGSLVLIILICSVFAVVT 299
Db 237 KSARVGWPACTSEC--NKEEYANLIDLIDGKTSGQTCNVHGIACLIIFVNLAFIGSLAV 294
```

```
Qy 300 ILLVGF-----RVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFF 341
Db 295 LLCVGFVFTSYDEDYTNLRQTSGESPKENRIK-----RTGAWI-----HNFENNA 342
Qy 342 QGWGTWVASWPUITLIVLSVIPVVALAAGLVFTLTDTPVELWSAPNSQARSEKAFDQHF 401
Db 343 RDIGMAGRNPKSHFFIGCAVLIFCLPGMIYHKESTNVVDMSSPRSRARQEOMVFNANF 402
Qy 402 GPFRTNQVILTAPNRSSVRYDLSLLGPKNPGSGLDLDLLLELLEQLERHLHQVWSPEA 461
Db 403 GRPQRYQQIMLL--SHRDFQSSGKLYG-----VFHKDIFEBELFDILNAIKNISTQSDG 455
Qy 462 QNISIQDICYAPLNPNDNTSLYDCCINSLLQFQNNRTLLLTANQT-----L 509
Db 456 -RTIILDDVCYRPMGPG-----YDCLIMSPNTYFQGNKEHDMKSNKEETVSEDDDAFDYF 510
Qy 510 MGQTSQVDWKHFLYCANAPLTFKDGATLALSCMADYGAIPVFPFLAIGGYKGDYSEABA 569
Db 511 SSEATTDEWMNHMAACIDQPMQSK--TKSGLSCMGTYGGPSAPNM-VFKNSTNHQAANS 567
Qy 570 LIMTESLANNYPAGDPRLAQAQKLWEAFLEEMPAFORRMAGMFQVTFTAERSLEDEINRT 629
Db 568 IMMTILVTQ--RTEPEIQKAELEWEKFLKFCKEYREKSPKVI-FSFMASRSITDEIENDA 624
Qy 630 AEDLPFIPATSYIVIFLYISLALGSY-----SSWSRVMYDVS KATLGLGGVAVVLGAVMAAM 684
Db 625 KDEIVTVVIALAFLIGYVTFSLGRYFCVENQLWS-ILVHSRICLGMLSVINLLSGFCSW 683
Qy 685 GFFSYLGRSSRLVLQVVPFLVLSVGADNIFIFVLEY--QRLP--RRPGEPREVHIGRA 739
Db 684 GIFSMFGIHPKVALVQFVVVTLGVCRTFMVVKYAYAQQRVSMYPMSPDQCEI-VGMV 742
Qy 740 LGRVAPSMLLCSLSEAICFFLGPALTPMPAVRTPALTSGLAVILDLELLQMSAFVALLSDS 799
Db 743 MAGTMPAMFSSSLGCASFIFGTFGTDLPAIRTFCLYAGLVIDVLVHTCTIFLALFVMDT 802
Qy 800 KRQEASRLDVCCVKPQELPPPQOGEGL---LLG-----FPQKAVAPELLHWI 844
Db 803 QRELNG-----KP-EFFPYQIKDLLGAYLIGRQRATDTFTMTQFFHFQVAPFLMHRM 853
Qy 845 TRGVVLLFLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEYGAIFYFT 904
Db 854 TRIITGIIFIASFITTVILSSKISVGPDSQMAFTEKSYISTHFRYLDKDFDVGPPVFFTV 913
Qy 905 TLGNFSEAGMNAICSSAGCNNSFTQKIOYATEFPPEQSYLAIPASSWVDDFDLWLT-P 963
Db 914 DGEILDWRHPDVQNKRFCTPPGCSDTSGFNIMNYAVGHTETQYLSGEMYNWIDNLEWISRK 973
Qy 964 SSCRLYISGPNKDKFCPSTVNSL---NCLKNCM-----SITMGSV--RPSVEQPH 1009
Db 974 SPCKVYVHDDN--TFCTNRNKSALDDKACRTCMDFDYVANSYPKSSIMYHRPSIEVFY 1031
Qy 1010 KYLPWFNLDRPNIKCPKGLAAYSTSVNLTSDGVQLASRFMAYHKPL--KNSQDYTEALR 1067
Db 1032 RHLRHFLDEDTNSCVCFGRASFKDAISFTSGRIQASQFMTFHKLSISNSSDFIKAMD 1091
Qy 1068 AARELANITADLRKVPGTDPAFEVFPYTIINVFYEQYLITLPEGLFMLSCLVPTFAVS 1127
Db 1092 TARMVSRRLERSI-----DDTAHVFPAYSKIFPFPEYQYSTIMPLITLTQLFTVVGVFGII 1145
Qy 1128 CLILGLDLRLSGLLNLLSVMILVDTVGFPMALWDISYNVLSINLVSAVMSVVFVSHITR 1187
Db 1146 CVTLIGIDVKGACAVICQVSNYFHVAFMIFNIPVNLASATNLVMSGDIIEFSVNLK 1205
Qy 1188 SFAISTKPTMLERAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQLIQIIPFFRLNLLIT 1247
Db 1206 GYACSLQRADRAESTVSGIPILSGPVVTVAGSTWMTFSGAHLQIITVVFFKFLITI 1265
Qy 1248 LLGLLHGLVFLPVLVSYGPDVNPALALEOKRAEAVAANVAVASCPNHPRSVSTADNIYV 1307
Db 1266 VSSAVHALIITLIPILLAFGSGRHSSTSTNDNEQHDACVLS--PTAESHISNVESGIL 1323
Qy 1308 N-----HSFEFSIKGAGAIISNLFNNGRQF 1332
```



[illegible]

Db	862	GSDGVLAYKLLVOTGSRDK--PIDISQLT--KORLVDADGIINPSA--FYIYLTAWVSN	915
Qy	1018	D-----RNIKCPKGGAAAY--STSVNLTSDGQVLASRFMAYHKPLKNSQD	1061
Db	916	DPVAYAASQANIRHREFWHDK---ADYMPETLRIAPAAEPTEYAQFPYFLNGLRDTSD	972
Qy	1062	YTEALRAARELANITA-DLRKVPGTDPDAPEVPPYITINNVYQYITILPEGLFMLSLCL	1120
Db	973	FVEAIEKRVVICNNYISLGLSSYPNG-----YFP---LFWBQYISLRHWLLLSISVVL	1022
Qy	1121	VPTFAVCSLLGLDLRSLGILLNLLSIYVMLVDTVGFNALWDISYNAVSLINLISAVGMSVE	1180
Db	1023	ACTFLVCVAFELLPWTFAGII--VWVLAMTVLEFGMGLIGIKULSAVPVLLIASVGIGVE	1081
Qy	1181	FVSHITRSP--AISTKPTWLERAKEATISMGSAVFAGVAMTNLPGLIVLGLAKAQIUIQIF	1238
Db	1082	FTVHVALAFLTAIGDKN---HRAWLALEHMFAPVLDS-AVSTLLGVLMLAGSEFDFIVRY	1137
Qy	1239	FFRLNLLITLLGLLHGLVFLPVILSYVG--PDVNPALAEQ---KRAERAVAAVMVASCP	1293
Db	1138	FFVAIAILTVLGVNLGLVLLPVLLSPFGPCPEVSPANGLNRLPTSPPEPPPSVVRFAVPP	1197
Qy	1294	NHPSRYS--TADNTYVNVHSPGSI-----KGAG	1319
Db	1198	GHTNNGSDSDSEYSQTTVSGISELRQYEAQQGAG	1234

RESULT 6

Tl8538  
 patched protein - chicken  
 C/Species: Gallus gallus (chicken)  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C/Accession: Tl8538  
 R/Marigo, V.; Scott, M.P.; Johnson, R.L.; Goodrich, L.V.; Tabin, C.  
 Development 122, 1225-1233, 1996  
 A>Title: Conservation in hedgehog signaling; induction of a chicken patched homolog  
 A/Reference number: Z18958; MUID:96205046; PMID:8620849  
 A/Accession: Tl8538  
 A/Status: preliminary;  
 A/Molecule type: mRNA  
 A/Residues: 1-1442 <MAR>  
 A/Cross-references: UNIPROT:Q90693; UNIPARC:UPI00001328B4; EMBL:U40074; NID:g1333  
 C/Genetics:  
 A/Gene: PTC  
 C/Superfamily: Drosophila membrane protein patched  
 C/Keywords: transmembrane protein

Query Match	9.1%;	Score 628.5;	DB 2;	Length 1442;	
Best Local Similarity	23.6%;	Pred. No. 1.6e-35;			
Matches	284;	Conservative 165;	Mismatches 421;	Indels 335;	Gaps 47;
Qy	365	ALAAGLVFTLTTDPVELWSA	NSQARSEKAFH	DQHFPPRTN-QVIL	TAPNRSSRYD 423
Db	111	AFAVGLRAANLETNVEELW	VEGVRSRELNYTRQ	KIGSEAMFNQ	LMIQTPQBDG--- 166
Qy	424	SLLLGPNFSGILDLLLELE	-LQERLRLQVWS	PEAQNTISLQ	ICVAP----- 474
Db	167	-----TNVLTTEALRQ	HLDSALQASRVH	VMYN---ROWK	LEHLCYKSGELTEA 213
Qy	475	--LNPNTSLYPCCINSLLQ	VFQNNRTLLLTANQ	LMGQ-----	512
Db	214	GYNDQIIIEVYLCILITP	LDPCWEGAKIQSGTA	--YLLGKPPLO	WINFDPLEFLBELKXI 271
Qy	513	TSQVWDKDFLY-----	CAN-----APLTF	--KDG	T-----ALALS-CMADY 546
Db	272	NQVQSVSWEEMLNKA	VEGHCHYMDRPPCLN	PADPCPITAPN	KNSTKPLDVALVLSGC---Y 328
Qy	547	GAPVFPF-----LAIGG	---YKGDYGEA	LIMTFSL-----NNYP	----- 580
Db	329	GLSRXYMHMQEBELI	GGTVKXSSGKLVS-AQ	ALQTMFLQMT	PKQMYEHFKYEVVSHINW 387
Qy	581	ACGPRLAQAKUWE	EAFLEBEMARFQRM	AGMFGQVTTFA	ERSLEDSINRTABDLPIFATSY 640

Mon Apr 10 07:16:16 2006

388 NEDKAAALEAWQRMVYEVVHOSVAQNSTQKVLSTTT--TTLDDILKSDSDYVIRVASGY 446  
 Db  
 641 IVIFLYISIALGYSWSRVMDSKATLGLGVAVVVLGAVMAAGFFSYGLRSLSLVILQ 700  
 QY  
 447 LLMLAYACITM---LRWD--CAKSOQAVGLAGVLLVALSVAAGLGLSLIGISFNAATQ 501  
 Db  
 701 VVPEFLVSGADNIFIVLEYORLPRRPGEPREHVGIRALGRVAPSMMLCSLSEALCEPL 760  
 QY  
 502 VLPEFLAGVGVDDVFLAHAFSETGQNKRIPEEDRTGECLKRTGASVALTSISNVTAFFM 561  
 Db  
 761 GALTMPAVRTFALTSLGLAVILDFLLQMSAFVALLSLDSKQASRLDVCCLC----- 812  
 QY  
 562 AALIPALRAFSLQAQVVVVFAMVLLIFPAILSMDLVRRDRDLDFCCFTSPCVTR 621  
 Db  
 813 ---VKPQ-----ELPPGGGEGE----- 827  
 QY  
 622 VIQEPQAYAEADNICVSSPPPYSSHFAHETQITMQSTVOLRTEYDPHTQAYYTTAEPR 681  
 Db  
 828 ---LLGFEQKAYAP 838  
 QY  
 682 SEISVQVPTVQDSLSCSPESASTRDLLSQFSDSVHCLPEPCTKWTLLSTFAEKHYAP 741  
 Db  
 839 FLLHWITRGVLLFLALFGLVSLYSMCHI SVGLDOELALPKDSYLLDYFLFLNRYFEVGA 898  
 QY  
 742 FLLPKAKAVVIFLFLGLLGLSLYGTTRVRDGLDLDIVPRDITREYDFIAAQKYFSF-- 799  
 Db  
 899 PVYFTTLGYNFSSEAGNAICSSAGCNFSGTQKIQYAT-----EFPEQSILAI 948  
 QY  
 800 ---YNN-----YIVTQADYPNVQHLLYELHRSFNSVTVLL 833  
 Db  
 949 PA-----SSWDDDFIDWL-----TPSS-----CRLYISGPNKD 977  
 QY  
 834 EGDRLPKMWLHYFRDLQGLQDAFDSWETGKITYSNYKNGSDDAVLAYLLVOTGNRA 893  
 Db  
 978 KFCPSVTNSLNCNKMSITMGSVRPSVEQPHKYL-PWFLND-----RPN1 1022  
 QY  
 894 K-PIDISQLT--KORLVDADGIINPNA--FYIYLTAWNSNDPVAAYASQANIRPHRPEW 947  
 Db  
 1023 KCPKGGLAAY--STSVNLSDGOVLASREFMAYHKPLKNSQDYTEALRAARELAANITA-D 1079  
 QY  
 948 VHDK---ADYMPETRURIPAAPIEVAQFPFYINGLRETSDFVEAIEKRAICNNYTSIG 1004  
 Db  
 1080 LRKVPCTDPAFEVFPYTIINVFYEQYLTILPEGLFMLSCLIVPTFAVSCLLGLDLRSL 1139  
 QY  
 1005 IASYPNG-----YFP-----LPWEQVIGLRHMLLSISVVLACTFLVLCALFLNPTAGI 1054  
 Db  
 1140 LNLISVIMILVDTVGFMAWDISYNAVSLINLSAVGMSVEFVSHITRSF--AISTKPTW 1197  
 QY  
 1055 I-VVVLALMTVELFGMGLIGIKLSAVPVVILLIASVGIVGFEFTVHALAFLTAIGDKN-- 1111  
 Db  
 1198 LERAKEATISMGSAVAVGAVMTNLPGLVILGLAKAQIQTFFRLNLITLLGLLHGLVF 1257  
 QY  
 1112 -RRAVTALAHMFAVLDO-AVSTLLGVLMAGSEDFIVRYFFAVLAILITLGVNLGLVL 1169  
 Db  
 1258 LPVLISVVG--PDVNPALA---LEQKRAEAEAAVAVASCPNHPSPRVSTADNIVNHSPE 1312  
 QY  
 1170 LPVLISFGPYPEVSPACGRNRLTPSPPEPPSIVRFALPPGHTNGS--DSSDSSEYSSQ 1227  
 Db  
 1313 GSIKG 1317  
 QY  
 1228 TVVSG 1232  
 Db

RESULT 7  
 S44797  
 F09G8.4 protein - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
 C:Accession: S44797  
 R:Anderson, K.  
 submitted to the EMBL Data Library, February 1993  
 A:Description: Sequence of the C. elegans cosmid F09G8.  
 A:Reference number: S44776

A:Accession: S44797  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-915 <AND>  
 A:Cross-references: UNIPROT:P34389; UNIPARC:UPI000013B859; EMBL:L11247; NID:G156280; PII  
 C:Genetics:  
 A:Introns: 51/1; 214/3; 255/3; 326/3; 382/2; 409/3; 455/3; 505/3; 569/2; 702/3; 836/3  
 Query Match 8.7%; Score 600; DB 2; Length 915;  
 Best Local Similarity 22.7%; Pred. No. 8.9e-34;  
 Matches 236; Conservative 158; Mismatches 42; Indels 224; Gaps 37;  
 QY 10 LLWALLLRLAQSEPTTHQPCYCAFYDECNGNPNLSGLSLMTLSNVSCLSNTPARKITGD 69  
 Db 14 VLFLLLIHLAQ-----AKVM--TECDGEEDSNHPCKNTKSTVLPITVTRSLNPT 64  
 QY 70 HLILQKICPLRYTGN--TQACCSAKQLVSLSEASITKALLTRCPACSDNPNVNLCHNT 128  
 Db 65 YNARFEKYSYLVQEDKAQVCCTELQKGMTRISNAATILGSCPCDFNFAKWCQFT 124  
 QY 129 CSPNQSLFINVTRVAQLGAGQLPAVYAEAFYCHSFAEQSYDSCSRVRVPAATAATLAVGTM 188  
 Db 125 CSPDQSKFMKVMET-----TGPKNVVVKMEFKVNRDFEGLYESCRHTWANGLAFLMSL 180  
 QY 189 CGVGSALCNAQRLNFGDGTG-NGLA---PLDITPHLLEPGQAVGSGIOPLMNEGVARN 244  
 Db 181 GGRVS-----FENFYGMGTGNLAQSIPIINTEFQFSRMKNAMNIPITP-----CH 225  
 QY 245 ESQDDVATCSCQDC---AASCPAIAARPOALDSTFYLGQMPGSLVLIILILCSVPVVUTIL 301  
 Db 226 KSAQPKVPACAGTDCPTNAHQLVDISKVEHLGTVKVFHPHP--DFEWLLKICGCLIA-LTVL 283  
 QY 302 LVGF-----RVAPARDKSNVDPKGTSLSKLSFSTHTLLGQFQG-----WGTWVAS 350  
 Db 284 LVFTLKYSCHRRGAPNGEDGCVYDLGKGN-----LEVQEGLCARVANAVIK 330  
 QY 351 WPLTILVSLVPPVVALAAG-LVFTLTTDPVWELMSAPNSQARSEKAF-HQHQFGPPFRTN 408  
 Db 331 HPLIFVSLGLIVAAACCSGNFKFSLTHSDVQSAADGETRNEKKTIHS--FGPNHRIE 388  
 QY 409 QVILTAPNRSRYDLSLLGPKNFSGILDLDLLELLEQLERLHLQVWSEARNTSLQ 468  
 Db 389 QIFNLPTT-----KSMFNMLFEMFQLVGNIQNLT--ACYGNSSVKLD 432  
 QY 469 DICVAPLNPDNTSLVDCINSLLOFYQNNRLLLTLLTANQTLMGOTSDQVWDKDFLYCAN 528  
 Db 433 DICYKPIGKN-----HGCIMSPNTNYFQ----- 455  
 QY 529 PLTFKDGITALSCMADYGA PVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQ 588  
 Db 456 -----YTTARTMITVLING-----PE-DQ 474  
 QY 589 AKLWBEAFLEEMRAFORRMAGMFOVTTAERSLEDEINFTTAED--LPFATSYIVIFLY 646  
 Db 475 ATAWETAFLNMSRYEMKHANF---TFMTETSAEEIHTAVETDKITVSVIACNAVLIWI 531  
 QY 647 ISIALGSY--SSWSRVWVDSKATLGLGVAVVVLGAVMAAGFPYSLGIRSSILVQVWP 704  
 Db 532 TMLGINHPRESSILSALVHHKLLISISAVMISVISVCSIGMPSLFGVHATDNAVIVLFF 591  
 QY 705 LVLSVGADNIFIFVLEYQRLPRRPGEP-----REVH--IGRALGRVAPSMMLCSLSEAI 758  
 Db 592 VITCLGINRIFVIRTFQANGCHYCYGLPNISYREMNHRISNVNRRSIPVLVTLNLSICSTCL 651  
 QY 759 FL-GALTP-----MPAVRTFALTSGLAVIDLFLQMSAFVALLSLDSKQO----- 802  
 Db 652 FLAGGVLPYVSVSMPAVEVEARHAGLAIMLDTAFYLLVMLPLFOYDARREMSGKEIWPW 711  
 QY 803 -----EASRLDVCCCKQPELPPQCGEGLLGFPQKAYAPFLHWTIRGVVLLFLALFG 858  
 Db 712 YELSNESKINLCMEAVDGNLRSP-----VDWFKLAIAPIALLKKICRIWIAITFFVSLI 764  
 QY 859 VSLYSMCHISVGLDQELALP-----KDSYLLDYFLFLNRYFEVGA PVYFTV 904

```
Db 765 IACVCTLCLEFGNQWAFSEVCIFIFSHLNGSQTSYLTGKHQNMENLNIGPPLWV 824
Qy 905 TLGYNFSSEAGMAICSSAGCNFSFTQKIQ---YATEFPPEQSYLAIPASSWVDDFDWL 961
Db 825 EGDVKWHDPRKQNKFCITLACDDNSMGKIRSLAYAEY-KGNVYLHGDVNIWLDLSYLP 883
Qy 962 TP-SSCCRLYISGPNKDKFC 980
Db 884 HPRGSCCKM-----DGKQFC 898

RESULT 8
T25600
hypothetical protein C32E8.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25600
R;Gattung, S.
submitted to the EMBL Data Library, February 1997
A;Description: The sequence of C. elegans cosmid C32E8.
A;Reference number: Z20056
A;Accession: T25600
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-933 <GAT>
A;Cross-references: UNIPROT:P91129; UNIPARC:UPI00000611CF; EMBL:U88308; PIDN:AAB42325.1
A;Experimental source: strain Bristol N2; clone C32E8
C;Genetics:
A;Gene: CESP:C32E8.8
A;Map position: 1
A;Introns: 145/1; 177/2; 385/1; 418/3; 833/3; 878/3

Query Match 8.7%; Score 598.5; DB 2; Length 933;
Best Local Similarity 22.9%; Pred. No. 1.2e-33;
Matches 251; Conservative 181; Mismatches 362; Indels 303; Gaps 37;

Qy 341 FQGWGTWASWPIITLVSLVSVVVALAAGLVFTLTDPVELWSAPNSQARSK-AFH-- 397
Db 14 FROLGLICDHPLPFFVFPFLFTAAMGVGLHLNPLSDAVYLFPLGAQSKMERMSTHEK 73
Qy 398 ----DQHFGP---FFRTNQVILTAPNRSRYSDSLGPKNFGSLDLDLLELLEQER 450
Db 74 WPLTDNNYIPGRAVTSREIQVFTALARN-----DSNILDPKFANAVYQLDKY-----IQTR 124
Qy 451 LRHLQVMSPEAQRNISLQDICYAPLPNDNTSLYDCCINSLLQYFQNNRTLLLLLTANQTL 510
Db 125 VRVLH----- 129
Qy 511 GQTSQVDWKDHFLYCANAPLTFKDGTA-----LSCWADYG-----APVFPFLAIGY- 559
Db 130 -----NGHYYSYKNLCLOYKNGGCPNSKHVHILSDLNHHGNITYPVYFRFGSGGYI 181
Qy 560 -----KGXDYS-----AALMTSLNYP-----AGDPLAQAKLWEAFLEE 599
Db 182 GSSLGGVTVMKGENETDILASAKAWPMIYHLKPHPEMSYISGE-----WE-----LEL 230
Qy 600 MRAFORMAGMF-QVTFTRASLEDEINRTTASDLPIFATSYIVIFLYISLALGSY--SS 656
Db 231 GRMLTQVPEDPYISITYFHSQTLADELKNWADTLIPFIISITLLIVFSLCSLSFDGS 290
Qy 657 WSRVMVDSKATLGGGVAVVLGAVMAAMGFYSYIGIRSSILVILQVVPFVLVSGADNIFI 716
Db 291 FSIDVLSKPILSILGWSAGIALLTGVSFLSLMGMPYN-DIVGVMPFLVLAAGVDNMF 349
Qy 717 FVLEYQRLPRRCPGEPREHV--IGRAGRVAPSMLLCSLSEAICFFIGALTMPAVRTFAL 774
Db 350 MVAAV-----RRTSRTHVHERMGECLADAASVILITSSTDVLSFGVGAITTIPAVQICV 405
Qy 775 TSLGAVILDFLQMSAFVALLSDSKRQEAR-----LDVCCCVK----- 814
Db 406 YTGVAIFFAFYQITTFACALAMKHEASGRNSLFLIEAVSAEKRTSISTFORLNLGS 465
```

```
Qy 815 -POELPPPGQEGELLGFFOKAYAPFLHWTIRGWVLLLF-LALFGVSLYSMCHISVGLD 872
Db 466 VPDSASHDVKQPLTSRFFGFEWAPVLMHPVVRGIAMVWFVYLLGAS-YGCSRIKEGLE 524
Qy 873 QELALPKDSYLLDYFLFLNRYF-EVGAPVYFV-----TTLYGNFSSE 913
Db 525 FVNLLEDSDYAIHYRLLEKYFWKYGOQVQVINNAPDLRNHTSRDRVHAMVLDFAYSKH 584
Qy 914 A-GMNAICSSAGCNFSFTQKIQVATEFPPEQ-----SYLAIPASSWVDDFDWL 961
Db 585 AIGMESV-----QFWLFEMERYQKELEFQIIDDSPFYGLLHFLASKTNPLAEDIYW- 637
Qy 962 TPSSCCRLYISGPNKDKFCPSTVNSLNLCKMCSITMGSVRPSVEQPHKYLPMFLNDRPN 1021
Db 638 -----GMPD----- 642
Qy 1022 IKCPKGLAAYSIVNLTSQGVLAS-RFMAYHKPLKNSQDYTEALPAARELANITADL 1080
Db 643 -----DDNGTVMKSPFRFLGKMKDLVTTMDQTDATMSPREVAARM--- 681
Qy 1081 RKVPGTDPAPEVFPYITITNVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRGLL 1140
Db 682 -----PEFNVTFPPIWMFTDQYIIIPNTYQNIILVMIIVIAVLFIPOPMCS-LW 733
Qy 1141 NLLSIVMLVDVTGFMALNDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200
Db 734 VALACASIDFGVIGYMTLWGNLDAISMITIIMSIGFSDVSAHIAIYGVVSRDRTAAGR 793
Qy 1201 AKEATISMSGSAVPAVAMTNLPGILVLGLAKAQLQIIFPRLMLLITLLGLLHGLVLPV 1260
Db 794 VKEALSALGWPLSQG-AMSTIIIAVSLADIPAYMI-VTFPEKTVLSISLGLHGLVLPV 851
Qy 1261 ILSYV-----GPDVNPALALEQKRAEEA-----VAAVMVASCPNHPRSVS 1300
Db 852 LLISIFVRGCCITPSSPHGHPSAQKIEQIRIAAISSPDLRLTVAPLRASSPISFPFHRLE 911
Qy 1301 -TADNIYVNHSPFGSIK 1316
Db 912 YTDESPTVHNRSKNSIK 928

RESULT 9
T13952
membrane protein ptch2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T13952
R;Motoyama, J.; Takabatake, T.; Takeshima, K.; Hui, C.
A;Title: Ptch2, a second mouse Patched gene is co-expressed with Sonic hedgehog.
A;Reference number: Z17830; MUID:98122566; PMID:9462734
A;Accession: T13952
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1182 <MOT>
A;Cross-references: UNIPROT:Q35595; UNIPARC:UPI0000001587; EMBL:AB010833; PIDN:BAA2469
A;Experimental source: strain BALB/cCrSlc
C;Genetics:
A;Gene: ptch2
C;Superfamily: Drosophila membrane protein patched
C;Keywords: transmembrane protein

Query Match 8.4%; Score 583.5; DB 2; Length 1182;
Best Local Similarity 24.2%; Pred. No. 1.8e-32;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

Qy 330 SFSTHTLLGQ-----FFQG-----WGTWASWPLTILVLSVFPVVALAAGLVFTLT 376
Db 20 SSAPHILAGSLQAPLWLRAYFQGLLFLSLGCRIQKHGKVLFLGLVAFGALGLRVAVIE 79
Qy 377 TDPVELWSAPNSQARSKAFHDQHFG-PFFRTNQVILTAPNRSRYSDSLGPKNFSGI 435
Db 80 TDLEQLWEVGRSVQELHYTKELGEEAAYTSQMLI-----QTAHQEGGNVLTPE----A 131
```



QY	436	LDLDLLELLELQERLRLHQVWSPAQRNLSLDICY	---APLNPDN-----TSLYDCC	486	
Db	132	LD-----LHQAALTASKQVSLYKGSWDLNKCYKSGVPLIENGMIERMIKLPFCV	184		
QY	487	INSLLOYFQNNRTLLLTANQTLMGOTSOVDKD	--HFLYCANNAPLT-----FKDGTALA	539	
Db	185	ILTPLCDCFEGAK	---LOGGSAYLPGRPDIOWTNLDPOQLLEELGPFASLEGFRELDDKA	241	
QY	540	LSCMADYGAPVF	-----PFLA---IGGYKGDYS-----	565	
Db	242	QVGQAVGRCLDPDPDPCHPCPPSNAPNHSRQAPNVAQELSGGCHGFSHKFMHWQBELLGG	301		
QY	566	-----EAPALIMTP-----	SLNNYPAGDPPLAQAOKLWEAFLEEMKAFORMMAG	609	
Db	302	TARDLQQLIRAEALQSTFLMLSPROYEHRFGDYQTHDIGWSEEQASVMVAQWRRFVQ	361		
QY	610	MFQVTPFAESLE	-----DEINRTTAEDLPIFATSVIV--IFLYISLALGSYSWS	658	
Db	362	LAQEALPANASQQIHAFSSTLDDIIRASE	-----VSTRVVGYYLLMAYACVTMLRWD	417	
QY	659	RVMVDSKATLGLGGVAVVLGAVNMAAGFFSYLGRISLVLQVPPVLVSVGADNIFPV	718		
Db	418	--CAQSGAVGLAGVLLVALAVASGLGCLALLGITENAAATQVLPPLALGIGVDDIFLLA	475		
QY	719	LEYQRLPRRGPGEVPHI	GRALGVAPSMLLCSLSEAI	CFEFGALGTPMAVTRFALTSLG	778
Db	476	HAFTKAP	--PDTPLPERMGCLSTGTSVALTSVNNNVAFFMVALVPIPALRAFSLOAAI	533	
QY	779	AVILDFELLQMSAFVALLSLSDKQASRLDVCCC	-----VKPQE-----	817	
Db	534	VVGCNFAAVMLVPAILSLDLRRHRQLDVLCCFSSPCSAQVIOMLPQELGDRAPVPGI	593		
QY	818	-----LPP-----	PG-----QCEG--	826	
Db	594	AHLTATVQAFTHCEASSQHVVTLLPQOAHLLSPASDPLGSELSPGSGTRDLLSQEETG	653		
QY	827	-----LLLGFFOKAVAPFLLHWITRGVWVLLFLALFGVSLYSMCHISVGLDQ	873		
Db	654	POAACRPLLCAHWTLAHFARYQFAPLLOTTAKALVLLFFGALLGLSLYGATLVQDGLAL	713		
QY	874	ELALPKDSYLLDYFLFNRYFEVGAPVYFVTTLGYNFS	--SEAGMNAICSSAGCNFSFTQ	932	
Db	714	TDVVPRGTKEHAFSLAQRLFYSL	--YEVALVTQGGFDYAHSORAL-----FDLHQ	761	
QY	933	KIQYATEPPEQSILAI PAS	-----SWDDF-----IDWLTPSSCCRLYISGPNK	976	
Db	762	RFSSU	-----KAVLPPPATQAPRTWLHYHYSWLGIOQAAFDQDASGRITCHSVRNGSED	816	
QY	977	D	-----KFCPSTVNSLNCNKCMSITM-----GSVRPSEVQFHKYLPWFNLNDRPNIKCPK	1026	
Db	817	GALAYKLLIQTGNAQEPDLFSQATTRKLVDKEGLIPP	--ELFVGLTVVWVSSDPL-----	869	
QY	1027	GGLAA	-----YST-----SVNLTSDQVLAISREMAHYHKLKNSDQYTEALR	1067	
Db	870	--GLAASQANFYPPPEWLHDKYDTGTEMIRIPAAQPLEFAQFPLHLGLOKTADFEVAIE	928		
QY	1068	AARELAANI	--TADLRKVPGTDPAFEVFPYITINVEYQYLTILPEGLFMLSCL--LVPTF	1124	
Db	929	GARAACTEAGQGVHAYSGSPF	-----LFWEQYGL--RRCFLLAVCILLVCTF	976	
QY	1125	AVSCILLGLDRLSGLLNLLSIVMLIVDTVGFMAIWDISYNAVSLINLVSAGMSVEFVSH	1184		
Db	977	LVYCALLLSPTWAGLI	--VLVLAMTMVELFGMFLGKLSAIPVIVLVSIGIGVEFVTH	1035	
QY	1185	ITRSFAISTKPTWLERAKEATISMGSAVFAGV	--AMTNLPGLIVLGLAKAQLIQIFFFR	1241	
Db	1036	VALGFLTSGSNRLRAA	-----SALEQTFAPVTDGAVSTLGLLWLAGSNPDFFIIRFFV	1090	
QY	1242	LNLLITLLGLHLGVFLPVILSYVGP	1267		
Db	1091	VLTVTLGLLGLLGLLPLVLSILGP	1116		

QY	436	LDLDLLELLELQERLRLHQVWSPAQRNLSLDICY	---APLNPDN-----TSLYDCC	486	
Db	132	LD-----LHQAALTASKQVSLYKGSWDLNKCYKSGVPLIENGMIERMIKLPFCV	184		
QY	487	INSLLOYFQNNRTLLLTANQTLMGOTSOVDKD	--HFLYCANNAPLT-----FKDGTALA	539	
Db	185	ILTPLCDCFEGAK	---LOGGSAYLPGRPDIOWTNLDPOQLLEELGPFASLEGFRELDDKA	241	
QY	540	LSCMADYGAPVF	-----PFLA---IGGYKGDYS-----	565	
Db	242	QVGQAVGRCLDPDPDPCHPCPPSNAPNHSRQAPNVAQELSGGCHGFSHKFMHWQBELLGG	301		
QY	566	-----EAPALIMTP-----	SLNNYPAGDPPLAQAOKLWEAFLEEMKAFORMMAG	609	
Db	302	TARDLQQLIRAEALQSTFLMLSPROYEHRFGDYQTHDIGWSEEQASVMVAQWRRFVQ	361		
QY	610	MFQVTPFAESLE	-----DEINRTTAEDLPIFATSVIV--IFLYISLALGSYSWS	658	
Db	362	LAQEALPANASQQIHAFSSTLDDIIRASE	-----VSTRVVGYYLLMAYACVTMLRWD	417	
QY	659	RVMVDSKATLGLGGVAVVLGAVNMAAGFFSYLGRISLVLQVPPVLVSVGADNIFPV	718		
Db	418	--CAQSGAVGLAGVLLVALAVASGLGCLALLGITENAAATQVLPPLALGIGVDDIFLLA	475		
QY	719	LEYQRLPRRGPGEVPHI	GRALGVAPSMLLCSLSEAI	CFEFGALGTPMAVTRFALTSLG	778
Db	476	HAFTKAP	--PDTPLPERMGCLSTGTSVALTSVNNNVAFFMVALVPIPALRAFSLOAAI	533	
QY	779	AVILDFELLQMSAFVALLSLSDKQASRLDVCCC	-----VKPQE-----	817	
Db	534	VVGCNFAAVMLVPAILSLDLRRHRQLDVLCCFSSPCSAQVIOMLPQELGDRAPVPGI	593		
QY	818	-----LPP-----	PG-----QCEG--	826	
Db	594	AHLTATVQAFTHCEASSQHVVTLLPQOAHLLSPASDPLGSELSPGSGTRDLLSQEETG	653		
QY	827	-----LLLGFFOKAVAPFLLHWITRGVWVLLFLALFGVSLYSMCHISVGLDQ	873		
Db	654	POAACRPLLCAHWTLAHFARYQFAPLLOTTAKALVLLFFGALLGLSLYGATLVQDGLAL	713		
QY	874	ELALPKDSYLLDYFLFNRYFEVGAPVYFVTTLGYNFS	--SEAGMNAICSSAGCNFSFTQ	932	
Db	714	TDVVPRGTKEHAFSLAQRLFYSL	--YEVALVTQGGFDYAHSORAL-----FDLHQ	761	
QY	933	KIQYATEPPEQSILAI PAS	-----SWDDF-----IDWLTPSSCCRLYISGPNK	976	
Db	762	RFSSU	-----KAVLPPPATQAPRTWLHYHYSWLGIOQAAFDQDASGRITCHSVRNGSED	816	
QY	977	D	-----KFCPSTVNSLNCNKCMSITM-----GSVRPSEVQFHKYLPWFNLNDRPNIKCPK	1026	
Db	817	GALAYKLLIQTGNAQEPDLFSQATTRKLVDKEGLIPP	--ELFVYGLTVVWYSSDPL-----	869	
QY	1027	GGLAA	-----YST-----SVNLTSDQVLA	RFEMAYHKLKNSDQYTEALR	1067
Db	870	--GLAASQANFYPPPEWLHDKYDTGTEMIRIPAAQPLEFAQFPLHLGLOKTADFEVAIE	928		
QY	1068	AARELAANI	--TADLRKVPGTDPAFEVFPYTTINVFYEQYLTILPEGLFMLSCL--LVPTF	1124	
Db	929	GARAACTEAGQGVHAYSGSPF	-----LFWEQYGL--RRCFLLAVCILLVCTF	976	
QY	1125	AVSCILLGLDRLSGLLNLLSIVMLIVDTVGFMAWLDISYNAVSLINLVSAGMSVEFVSH	1184		
Db	977	LVYCALLLSPTWAGLI	--VLVLAMTMVELFGMFLGKLSAIPVIVLVSIGIGVEFVTH	1035	
QY	1185	ITRSFAISTKPTWLERAKEATISMGSAVFAGV	--AMTNLPGLIVLGLAKAQLIQIFFFR	1241	
Db	1036	VALGFLTSGSNRLRAA	-----SALEQTFAPVTDGAVSTLGLLWLAGSNPDFFIIRFFV	1090	
QY	1242	LNLLTLTLGLHLVFLPVILSYVGP	1267		
Db	1091	VLTVTLTLGLHLGLLPLVILSILGP	1116		

```
QY 996 ITMGSVRPSVEQFHKYLPHFLNDRPNKCPKGGGLAAYSTSVNLTSQVQL-ASRFMAYHK 1054
Db 630 -----VFLSQPQF-----AKYNRDVVLTEDEGLEASRI----- 658
QY 1055 PLK-----NSQDYTEALRAARELANATADLRKVPGCTDPAFEVPPYITINVFQYKLTILP 1110
Db 659 PVQLRHVGANSQRAWRLFRRLAE--TSELQ-----TGVYADFFQ-----FAEQYNAVILP 706
QY 1111 EGLFMLSCLVPTFAVSCILLGLDLRSGLNLLSIVMILVDTVGFWMALWDISYNAVSLIN 1170
Db 707 GTLSSIAVAGVAVAVSLILPEPVAS-LWVSFVSINIGILGFMTFWSVRLDFISWT 765
QY 1171 LYSAVGMSVEFVSHITRSFAITKPTWLEAKETISMGSAVFAGVAMTNLPGIILVGLA 1230
Db 766 IVMSIGFCVDFAAHLAYNFAKQNMDSRERNALYAVGAPILMS-ATSTIIGVSFMASA 824
QY 1231 KAOLIQIPFRLNLTLLGLLGLVFLPVILS--VVGPDVNPALALQKRAEAAVAV- 1287
Db 825 ESYVFR-SFLKTIIMVILLAGLHGLVILPVLLSMFYCG-----SSKKAKEHIDAVD 875
QY 1288 -MVASCPNHPSRVST 1301
Db 876 QKIQAYNNPARTAS 890

RESULT 11
Ti8291
C:Species: Brachydanio rerio (zebra fish)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: Ti8291
R:Concordet, J.P.; Lewis, K.E.; Moore, J.W.; Goodrich, L.V.; Johnson, R.L.; Scott, M.P.;
Development 122, 2835-2846, 1996
A:Title: Spatial regulation of a zebrafish patched homologue reflects the roles of sonic
A:Reference number: Z18860; MUID:96379744; PMID:8787757
A:Accession: Ti8291
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1220 <CON>
A:Cross-references: UNIPROT:Q98864; UNIPARC:UIP0000132883; EMBL:X98883; PIDN:CAA67386.1
C:Genetics:
A:Gene: Ptc1
C:Superfamily: Drosophila membrane protein patched

Query Match 8.3%; Score 572; DB 2; Length 1220;
Best Local Similarity 22.4%; Pred. No. 1.2e-31;
Matches 270; Conservative 174; Mismatches 431; Indels 328; Gaps 43;

QY 321 KGTSLSDKLSFSTHTLLGQFPQGGTGWASWPLTILVLSVIVVVALAAGLVFTBLTDPV 380
Db 51 KGKAVGQKAPLWIRARFOAPFLSLGCHIQRHCGKVLFGLLVFGALSGLRVAAIETDIE 110
QY 381 ELASAPNSQARSEKAFHDQHFQ-PFRITNQVILTAPNRSSVRYDSLLGKPNFSGIILDD 439
Db 111 KLWVEAGSRVSKELRYTKERQGESVFTSQMLIQTPKQEG-----TNILTOE 157
QY 440 LLELELEQLRHLQVWSPSAQRNLSLODICYAPLNP--DNT-----SLYDCCINSL 490
Db 158 AL--LLHLEALASKQVSVLSYKGSWDLNKICFKSGVPIIENVMIERMIDKLPFCMIIVTP 215
QY 491 LOYFQNNRTL-----LLLTANQ--TLMGQTQSDVDWKDFLYCA 526
Db 216 LDFWEQSKLQGSAYLPGMPDIIQWMLDPLKMEELSQFTSLGPFREMLDKAQVGHAYM 275
QY 527 NAPTFFKDGKTALALSCWADYGP-----VPPFLAI-----GGYKG----- 561
Db 276 NRPECLDPSDT-----DC--PHSAENKPDQWQVNTAAELQGGCHGFSKFKFMHQBELILGER 329
QY 562 -KD-----YSAEALIMTFSL-----NNYPAGDPRLAQAKL-----WEEAPLEE 599
Db 330 VKDSQNALQAEALQTWFLMSKQLYEHFKDDYIEHDINWEDKATAILESQWRKFVEV 389
QY 600 MRAP--ORRMAGMFQVFTTAERSLEDEINRTAEDLPFATSVIVIFLYISLALGSYSSW 657
```

## RESULT 12

T29590

hypothetical protein F55F8.1 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T29590

R:Gattung, S.; Scheet, P.; Kemp, K.

submitted to the EMBL Data Library, November 1996

A:Description: The sequence of *C. elegans* cosmid F55F8.

A:Reference number: Z20647

A:Accession: T29590

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

```
Db 390 VHGSIPQSSNNVYAFSTI--TLNDIMKFSFSDVSVIRVAGGYLLMLAYACTVM---LRW 443
QY 658 SRVMVDSKATLGLGVAVVLGAVMAAMGFFSYLGRSSILVILQVVPFLVSLVGADNIFIF 717
Db 444 D--CAKSOQAGVAGVLLVALSVAAGLGLCSLLGLSFAAATTVQLPSLAGIGVDDMFL- 500
QY 718 VLEYQRLPRRCPCEPREVHIGRALGRVAPSMILCSSEALCPFLGALTMPAVRTFALTSG 777
Db 501 -LGHSFTTRSNIPKERTGDCRLRTGTSTVALTSVNNMIAFFMAALVPIPALRAFSLQAA 559
QY 778 LAVILDFLLQMSAFVALLSLDKROEASRLDVCCC-----VKPQE----- 817
Db 560 VVVVFNFAWALLIPALLSLDLHREDRLDILCCFYFSCSRVLIQIQOELSDANDNHQ 619
QY 818 -----LPPGQ----- 823
Db 620 RAPATPTVTGTTTTHITTTTVOAFTQDAAGQHIVTILPPTSQISTTPPSMVLSTPTP 679
QY 824 -----GEGE-----LLGFFQKAYAPFLHLHWTR 846
Db 680 TTDYGSQVFTTSSSTRDLLAQVEEPKEGECVPLPFRFNNLSSFAREKYAPLLKPKETK 739
QY 847 GVLLLLFLALFGVSLYSMCHISVGLDQELALPKDSYLDYFLFLNRYFEVCGAPVVFVTTL 906
Db 740 TVVVVVALLSLSLYGTTMVHDGLYLDIVPRDTQYEYEFITAQKYFSF-YNNMLVTVMD 798
QY 907 GYNFSSBAGMNAICSSAGCNFNFSPTQIQYATFEPQSYLA-----IPASSWVDDFDW 960
Db 799 GFDYAR-----SQQLQLHNAFNVSVKYVVKDGNHKLKLP-RMWLHYFQDW 841
QY 961 LTPSSCC-----RLYISGNPKKFCBPSTVNSLNLKNC 993
Db 842 LKGLQATPDADWEAGKITYSYRNGTBDGALAYKPLIQTSKGKGFNFYSQLTSSRLVDG- 900
QY 994 MSITWGSVRPSVEQFHKYLKLP-WFLNDRPNKCPKGGGLAA-----YST--- 1034
Db 901 ----DGLIPPEV--FYIYLTVWSND-----PLGYAASQANFYPHPREWHDKYDTTGE 948
QY 1035 SVNLTSQGVLASRFMAYHKPLKNSQDYTEALRAARELAN-ITADLRKVPGCTDPAFEVF 1093
Db 949 NLRIPAAEPLEFAQFPFLNGLRQASDFEIAESVRTICEBFMRQGIKNYPNG-----Y 1002
QY 1094 PYTITNVEYQYLTILPEGLFMLSCLVPTFAVSCILLGLDLRSGLNLLSIVMILVDTV 1153
Db 1003 PF-----LWFQYIQLRHFWFLLSISVWLACTFLVCAILLNPNWTAGVI-VFILPMWTVELF 1057
QY 1154 GFMALWDISYNAVSLINLVSAGVMSVEFVSHITRSF--AISTKPTWLEAKEATISGSA 1211
Db 1058 GIMGLIGIKLSAIPVVLIIASVGLIGVEFTVHIALGFLTAIGDRNT---RSVAMEHMFAP 1114
QY 1212 VFAGVAMTNLPGIILVGLAKAQILQIFPFRNLNLTLLGLLHGLVFLPVLSYVGP--DV 1269
Db 1115 VIDG-AISTLLGVMLAGSEDFIMRYFFAVLAILTLGLINGLVLPLVLLSLMGPPAEV 1173
QY 1270 NPA 1272
Db 1174 VPA 1176
```

A;Residues: 1-889 <GAT>  
A;Cross-references: UNIPROT:P91346; UNIPARC:UPI000017BA20; EMBL:U80447; PIDD:AA837812.1;  
A;Cross-references: UNIPROT:P91346; UNIPARC:UPI000017BA20; EMBL:U80447; PIDD:AA837812.1;  
A;Experimental source: strain Bristol N2; clone F55F8  
A;Genetics:  
A;Gene: CESP.F55F8.1  
A;Map position: 1  
A;Introns: 36/3; 66/3; 98/3; 149/2; 200/3; 240/2; 279/1; 358/1; 481/2; 569/3; 606/3; 697/3

	Query Match	7.8%;	Score 539.5;	DB 2;	Length 889;
	Best Local Similarity	23.0%;	Pred. No. 1.5e-29;		
	Matches 223;	Conservative 156;	Mismatches 337;	Indels	Gaps 35;
401 QY	FGPPFRNTQVIL-----TAPNRSRYRDSLLGPKFNGSGILDLDLLELE 446				
Db	31 FGPSYSTRRIIHDAPLVDGTFVAGRAVQS--REVQVAVVARSGNILDVFSNELKL 88				
QY	447 IQERLRHLQVWSPQAQRNISLQDICYAPLNPNDNTSLDYCCINSLLOQFQNNRTLLLTAN 506				
Db	89 MESFIRN-NITVQFSNRTWSPADLCLA--GPDGR-----CANNDHIQ----- 127				
QY	507 QTLMGQTSQVDWKDHFLYCANAPLTFKDGITALALSCADYGAVPVFPFLAIGGYK----- 560				
Db	128 --LASRLHQ-----HGINITYPTVRLSDKSAITIAS-----ALGGVKLAGDN 167				
QY	561 GKD-YSEAFALIMT:SLANNYP-----AGDPLQAQKLWEEAFLEHRAFORERWAGMFOV 613				
Db	168 GENIIIVEATWALLIYQLKFPNEISVSG--LWREFPKNDVEY-KKOAKYISI 218				
QY	614 TPTAERSLDEINRTTAEDLPATSYIVIFLY-----ISLALGSSYSSMRVWVDSKAT 667				
Db	219 TYFHSQTLSDELNRNAERLAPKFIGAFVILCVFSLCSIVIKSGYIDW-----VVTKPI 274				
QY	668 LGLGSHVAVLCAVMAWGFYSYLGRSSVILQVVPFLVLSVGADNIFIFVLEYQRLPR 727				
Db	275 LSVLGSVNAWGASAGMLTYLEIQYN-DIIAVMPFLVAVGTDNMFVASKRTRD- 332				
QY	728 PGEPREVHI GRALGRVAPSMMLCSLEACFFLGALTMPMPAVRTALTSGLAVIDFLQ 787				
Db	333 -NLKVDQRIACMADAASVILITALTALDSFGVTGTTIPAVQIFCIYTMCALLTFAYQ 391				
QY	788 MSAFVALLSLDSKEQ-----ASRLDV---CCVVPQBLPPPGOG----- 824				
Db	392 LTFFCAILVYTRYIEEQLHSIWLRPATVYSSTSLNKLFWLGSQPOK-PIPSCGTWSS 450				
QY	825 -----EGLLGHFKQKAYAPFLH-WITRGVLLFLALFGVSLYS 863				
Db	451 TSSVSTMTSQATSPASKHLHCAATSFPFNWYAPVLMQPMI-RAIAGLWLYILGISIYG 509				
QY	864 MCHI SVGLDOELALPKDSVLLDYFLFLNR-YFEVGAPVYVTT-----LGYNFSSE 913				
Db	510 CTHLKEGLEPANLLVDDSTATHYRVLEKHYWHYGASLQIVVSNPPDLRDFVERINMDKM 569				
QY	914 AGMNAICSSA-GGNNSF-----TQKIYATE-----FPQSIVLATPASSVDDDFI 958				
Db	570 ASTFANCKVAILGDSVQFWLREMQVSEIHKIQYDNKFPYDHAQAQIYSDMSQPVVDVV 629				
QY	959 DWLTPSSCKRLYISGNPKDKFCPSVTNSLNLCKNCSITMGSVRPSVEQFHYLPWFIND 1018				
Db	630 -----WGRNN 634				
QY	1019 RPNKCPKGLAAYSTSVNLTSQGVLAERFWAYHKPLKNSODYTEALRAARELAANITA 1078				
Db	635 -----NSERIITKFRFMIGMDISTTTKQTEATNTFREIASRF-- 672				
QY	1079 DLKRVGTDPAFVFPYTIINVEQVLTILPEGLFMLSCLCVPTFVNSCLLGLLDRSG 1138				
Db	673 -----EOYNVTTYMLPLFTDQVALVVPNTMQDIIIVAVACMLVISALLIPQVCSF 723				
QY	1139 LNLALSIVMLVDTVGFMAWMDI:SYNAVSLINLVSACMSVEFVSHITRSFAISTKPTWL 1198				
Db	724 WV-AVTIGSIDLVGLFWLNNVLDASIMTITMSVGFSDVSAHITYAVVISKESTTS 782				
QY	1199 ERAKEATTSMGSANFAGVAMTNLPGIILVGLAKAQLIQI:IFFFRNLNLITLLGLLHGLVFL 1258				



Db	857	MVTIDIGVIGLSLWSKLPDPISMITIISIGSIEFSAHITGFSVNSLNSAFRCVD	916	QY	522	FLYCANAPLTPKDGTAALASCM	---	ADYGAPVFPFLAIG--GYK	560
QY	1204	ATISMGSAVAGVAMTWLPGILVLGLAKAQILQIFFPRMLLTLGLLHGLVFLPVLIS	1263	Db	265	EQMKRAAIA--SGYMEKPCLNPNCPDTPAPNKNSTQPPDVGA	---	ILSGGCGYGA	317
Db	917	AMEKLAWPVVG-SLSIILGVFLAFIDSVMVLVFFKTIISLVL-IIGAMHALMLLPILLS	974	QY	561	GKDYSEAEALIMFTSLNNYPAGDPLAQ	---	---	589
QY	1264	YVGP-----DVPNPALALEQKRAEAAVAVMVASCP	1293	Db	318	AKHMHPEELIVGGAKRN-RSGHLRKAQALQSVVQLMTEKEMYDQWQDNKYVHHLGWTOE	---	376	
Db	975	MCIFVIERLSDASKKASDRRKLKNKENVVAINLP	1010	QY	590	-----KLWEAP--LEEMRAFQRMAGMFOVTTAERSLDEINRTTAEDLPIPATSY	---	640	
				Db	377	KAAEVLNAWQRFNSREVEQLLRKQRIATNYDIYVFSSAALDDILAKFSPHSALSIVIGV	---	436	
RESULT 15				QY	641	IVIFLYISLALGYSWSRVMDSKATLGLGGVAVVLGAVMAAMGFYSFIRSSVLILQ	---	700	
S06119				Db	437	NTVLVY--AFCTLLRM-RDPVGGSSVGVAGVLLMCFSTAAAGLGLSALLGIVFVNAASTQ	---	492	
membrane protein patched - fruit fly (Drosophila melanogaster)				QY	701	VVPFLVLSGANIFIVFLEYQRLPRRPGCEPREVHIGRALGVAPSMLLCSSEALCFEL	---	760	
C:Species: Drosophila melanogaster				Db	493	VVPFLALGLVDHIFMLTAAAYAESNR--EQTKL-----ILKKVGPSILFSACSTAGSFFA	---	546	
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004				QY	761	GALTMPAVRTFALTSLGLAVILDFLLQMSAFVALLSLDSKROEASRLDV-CCCV-----K	---	814	
C:Accession: S06119; A33468				Db	547	AAFIPVPAUKVFCQAATVWCSNLAALIVFPAMISLDLRRRTAGRADIFCCCFPVWKEQ	---	606	
C:Nakano, Y.; Guerrero, I.; Hidalgo, A.; Taylor, A.; Whittle, J.R.S.; Ingham, P.W.				QY	815	PQELPP-----PQGEGL-----LLGPF	---	832	
Nature 341, 508-513, 1989				Db	607	PKVAPPVLPLNNGNGRGARHPKSCNNRVVALPAQNPLLEQRADIPGSSHSLASFSLATFA	---	666	
A:Title: A protein with several possible membrane-spanning domains encoded by the Drosoph				QY	833	QKAYAPFLHWTGRVVVLLFLALFGVLSYMCCHISVGLDDELALPKDSYLLDYFLFLNR	---	892	
A:Reference number: S06119; MUID:90015164; PMID:2797178				Db	667	FQHYTFPLMSRWKFLTVMGFLAALISLYASTRLQDGLDIIDLVPKDSNEHFKFLDAQTR	---	726	
A:Accession: S06119				QY	893	YEEVGAPVYFVTTLGYNFSSEAGNNAICSSAGCNFSTQKIQYATATEFFPQSYLAIP---	---	949	
A:Status: not compared with conceptual translation				Db	727	LF--GFYSMTAVTQG-----NFEYPTQQQLLRDY-HDSFVRVPHVI	---	764	
A:Molecule type: DNA				QY	950	-----ASSWDDFDLWLPSSCCRLYISGPNKDFPCPSTVNSLCKNC-----	---	993	
A:Residues: 1-1299 <NAK>				Db	765	KNDNGGLPDPFWLLLFSEWL-----GNLQKIFDEEYRDGRLTKECWFPNASSDA	---	812	
A:Cross-references: UNIPROT:P18502; UNIPARC:UPI000014EA18; GB:X17558; NID:g8389; PIDN:CA				QY	994	-----MSITMGSVRPSVEQ-----FHKYL-PWFLNDRPNIKCPKG	---	1028	
R:Hooper, J.E.; Scott, M.P.				Db	813	ILAYKLIVQTHVDNPDVDELVTNRLVNSDGIINQRAFTNYLSAW-----ATNASSPTEL	---	868	
Cell 59, 751-765, 1989				QY	1029	LAA--YSTVNLTSDDGOVLASRFMAYHKPLKNSQD-----YTEALRAAR	---	1070	
A:Title: The Drosophila patched gene encodes a putative membrane protein required for se				Db	869	LRANCIRNRANGASQGLYPEPROYFHQP--NEYDLKIPKSLPLVYAQMPPYLHGLTDS	---	926	
A:Reference number: A33468; MUID:90058658; PMID:2582494				QY	1071	ELAANITADLRKVPGTDPAFEV--FPYTTINVFYEQHLTILPEGLFMLSCLVFTFAVSC	---	1128	
A:Accession: A33468				Db	927	QI-KTLIGHIRDLGVKVEGFLPNYPGIPPFIFWEQYMT-LRSSLAMILACVLLAALVIV	---	984	
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra				QY	1129	LLGLDLRSGLNLLLSIVMILVDTVGFMALWDISYNAVSLINLVSAGMSVEFVSHITPRS	---	1188	
A:Molecule type: mRNA				Db	985	SLLLSVMAAVLVILSVLSLAQAFGATLLGILKLSAIPAVILSVGMCLCF--NVLLIS	---	1042	
A:Residues: 1-110,'R',112-273,'G',275-331,'R',333-635,'P',637-861,'DVF',878,'Y',880-1299				QY	1189	FAISTKPTWLBRAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQILQIFFRLNLLITL	---	1248	
A:Cross-references: UNIPARC:UPI000016BCA3; GB:M28418; GB:M28999; NID:g552097; PID:g55209				Db	1043	LGFMITSVGNRRQRRVOLSQMQLSGPLVHGMLTSGVAVFMLSTSPPEFVIRHFECWLLLVLC	---	1102	
C:Genetics:				QY	1249	LGHLHGLVFLPVILSYGQPDVNPALALEQKRAEAAVAVMVASCPNHPSPRYSTADNIVYN	---	1308	
A:Gene: FlyBase:ptc				Db	1103	VGACNSLLVFPILLSVMVGPEAE-LVPLE-----HPDRISTPSPLVR	---	1143	
A:Cross-references: FlyBase:FBgn0003892				QY	1309	HS	1310		
A:Map position: 2 44D3-D4				Db	1144	SS	1145		
C:Superfamily: Drosophila membrane protein patched									
F:74-92/Domain: transmembrane #status predicted <TM01>									
F:427-448/Domain: transmembrane #status predicted <TM02>									
F:456-503/Domain: transmembrane #status predicted <TM03>									
F:529-555/Domain: transmembrane #status predicted <TM04>									
F:557-585/Domain: transmembrane #status predicted <TM05>									
F:677-699/Domain: transmembrane #status predicted <TM06>									
F:967-1017/Domain: transmembrane #status predicted <TM07>									
F:1019-1047/Domain: transmembrane #status predicted <TM08>									
F:1061-1086/Domain: transmembrane #status predicted <TM09>									
F:1093-1121/Domain: transmembrane #status predicted <TM10>									
F:142,298,335,388,807,861,1194,1271/Binding site: carbohydrate (Aen) (covalent) #status									
Query Match									
Best Local Similarity									
Matches									
321									
43									
381									
103									
437									
154									
487									
205									

## Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

**When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.**

**THIS PAGE IS BLANK**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:13:14 ; Search time 265 Seconds  
(without alignments)  
3546.278 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 MAEAGURGWLWALLRLLAQ.....GSIKAGATSNFLPNNGRQF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6872.5	99.5	1359	1 NPCL1_HUMAN	Q9uhc9 homo sapien
2	5421.5	78.5	1331	1 NPCL1_RAT	Q6t3u3 rattus norv
3	5407	78.3	1333	1 NPCL1_MOUSE	Q6t3u4 mus musculu
4	3179	46.0	1132	2 Q4T749_TETNG	Q4t749 tetraodon n
5	2413	34.9	1277	2 Q9JLG3_CRIGR	Q9jlg3 cricetulus
6	2412	34.9	1277	1 NPCL1_PIG	P56941 sus scrofa
7	2405	34.8	1276	2 Q8MKD8_FELCA	Q8mkd8 felis silve
8	2404.5	34.8	1276	2 Q9GK52_CANPA	Q9gk52 canis fami
9	2402.5	34.8	1278	1 NPCL1_HUMAN	O15118 homo sapien
10	2402	34.8	1289	2 Q59GR1_HUMAN	Q59gr1 homo sapien
11	2400	34.7	1277	2 Q7TMD4_MOUSE	Q7tmd4 mus musculu
12	2399	34.7	1276	2 Q9NQOQ_FELCA	Q9nqo0 felis silve
13	2395	34.7	1276	2 Q8MI49_FELCA	Q8mi49 felis silve
14	2385	34.5	1278	1 NPCL1_MOUSE	O35604 mus musculu
15	2367.5	34.3	1277	2 Q9GLC9_BOVIN	Q9glc9 bos taurus
16	2361	34.2	1286	2 Q9TT75_RABIT	Q9tt75 oryctolagus
17	2338	33.8	1209	2 Q4RWY5_TETNG	Q4rwy5 tetraodon n
18	2282	32.7	1287	2 Q9V124_DROME	Q9v124 drosophila
19	2256	32.7	1287	2 Q9U5W1_DROME	Q9u5w1 drosophila
20	2253	31.6	1287	2 Q7YU59_DROME	Q7yu59 drosophila
21	2205	31.9	1291	2 Q7Q409_ANOGA	Q7q409 anopheles g
22	1864.5	27.0	1233	2 Q9VRC9_DROME	Q9vrc9 drosophila
23	1760	25.0	1003	2 Q7PS03_ANOGA	Q7ps03 anopheles g
24	1725	25.0	1275	2 Q5LNK7_MAGGR	Q5lnk7 magnaporthe
25	1712.5	24.8	1361	2 Q7XUB7_ORYSA	Q7xub7 oryza sativ
26	1700.5	24.6	1271	2 Q5BBG1_EMENI	Q5bbg1 aspergillus
27	1659.5	24.0	1330	2 Q5KGS9_CRYNE	Q5kgs9 cryptococcu
28	1658.5	24.0	1330	2 Q5SSD4_CRYNE	Q5ssd4 cryptococcu
29	1643	23.8	1273	2 Q4WNG5_ASFPU	Q4wns5 aspergillus
30	1614.5	22.4	1295	2 Q410K4_GIBZE	Q410k4 gibberella
31	1549	22.4	1264	2 Q6BT03_DEBHA	Q6bt03 debaryomyce

32	1527.5	22.1	1275	2 Q9SHN9_ARATH	Q9shn9 arabidopsis
33	1511.5	21.9	1342	2 Q9TVK6_DICDI	Q9tvk6 dictyosteli
34	1508	21.8	1239	2 Q6CBAL_YARLI	Q6cbal yarrowia li
35	1500	21.7	1162	2 Q7RWL9_NEUCR	Q7rw19 neurospora
36	1461.5	21.2	1256	2 Q592V0_CANAL	Q592v0 candida alb
37	1452.5	21.0	1397	2 Q551C5_DICDI	Q551c5 dictyosteli
38	1437.5	20.8	1489	2 Q4PEB3_USTMA	Q4peb3 ustilago ma
39	1394	20.2	1055	2 Q9SVF0_ARATH	Q9svf0 arabidopsis
40	1386.5	20.1	620	2 Q4H344_CIOIN	Q4h344 ciona intes
41	1359	19.7	1382	2 Q618V4_CAEBR	Q618v4 caenorhabdi
42	1341.5	19.4	1383	1 NPCL1_CAEEL	Q19127 caenorhabdi
43	1329	19.2	1170	2 Q12200_YEAST	Q12200 saccharomyc
44	1323	19.1	1339	2 Q50RB6_ENTHI	Q50rb6 entamoeba h
45	1312.5	19.0	1178	2 Q750G1_ASHGO	Q750g1 ashbya goss

ALIGNMENTS

RESULT 1	NPCL1_HUMAN	STANDARD;	PRT;	1359 AA.
ID	NPCL1_HUMAN	Q9UHC9; Q6R3Q4; Q9UHC8;		
AC	Q9UHC9; Q6R3Q4; Q9UHC8;			
DT	13-SEP-2005 (Rel. 48, Created)			
DT	13-SEP-2005 (Rel. 48, Last sequence update)			
DT	13-SEP-2005 (Rel. 48, Last annotation update)			
DE	Niemann-Pick C1-like protein 1 precursor.			
GN	Name=NPCL1;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]_TaxID=9606;			
RP	NUCLEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 2 AND 3), AND TISSUE SPECIFICITY.			
RX	MEDLINE=20247253; PubMed=10783261; DOI=10.1006/geno.2000.6151;			
RA	Davies J.P., Levy B., Ioannou Y.A.;			
RA	"Evidence for a Niemann-Pick C (NPC) gene family: identification and characterization of NPC1L1";			
RL	Genomics 65:137-145(2000).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2), AND TISSUE SPECIFICITY.			
RX	PubMed=14976318; DOI=10.1126/science.1093131;			
RA	Altman S.W., Davis H.R. Jr., Zhu L.-J., Yao X., Hoos L.M.,			
RA	Tetzloff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,			
RA	Murgolo N., Graziano M.P.;			
RT	"Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol absorption.";			
RL	Science 303:1201-1204(2004).			
RN	[3]			
RP	TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.			
RX	PubMed=15671032; DOI=10.1074/jbc.M409110200;			
RA	Davies J.P., Scott C., Oishi K., Liapis A., Ioannou Y.A.;			
RT	"Inactivation of NPC1L1 causes multiple lipid transport defects and protects against diet-induced hypercholesterolemia.";			
RL	J. Biol. Chem. 280:12710-12720(2005).			
RN	[4]			
RP	INDUCTION.			
RX	PubMed=15604518; DOI=10.1194/jlr.M400400-JLR200;			
RA	van der Veen J.N., Kruit J.K., Havinga R., Baller J.F.W., Chimini G.,			
RA	Lestavel S., Staels B., Groot P.H.E., Groen A.K., Kuipers F.;			
RT	"Reduced cholesterol absorption upon PPARGdelta activation coincides with decreased intestinal expression of NPC1L1.";			
RL	J. Lipid Res. 46:526-534(2005).			
RN	[5]			
RP	FUNCTION.			
RX	PubMed=15928087; DOI=10.1073/pnas.0500269102;			
RA	Garcia-Calvo M., Lienock J., Bull H.G., Hawes B.E., Burnett D.A.,			
RA	Braun M.P., Crona J.H., Davis H.R. Jr., Dean D.C., Demers P.A.,			
RA	Graziano M.P., Hughes M., Macintyre D.E., Ogawa A., O'Neill K.A.,			
RA	Iyer S.P.N., Shevell D.B., Smith M.M., Tang Y.S., Makarewicz A.M.,			
RA	Ujjainwalla F., Altmann S.W., Chapman K.T., Thornberry N.A.;			

"The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1).";  
[6]  
PROG. Natl. Acad. Sci. U.S.A. 102:8132-8137(2005).  
VARIANTs LEU-55 AND ASN-1233.  
PubMed=15679830; DOI=10.1111/j.1399-0004.2004.00388.x;  
Wang J., Williams C.M., Hegele R.A.;  
"Compound heterozygosity for two non-synonymous polymorphisms in NPC1L1 in a non-responder to ezetimibe.";  
RT Clin. Genet. 67:175-177(2005).  
CL - FUNCTION: Play a major role in cholesterol homeostasis. Is critical for the uptake of cholesterol across the plasma membrane of the intestinal enterocyte. Is the direct molecular target of ezetimibe, a drug that inhibits cholesterol absorption. Lack of activity leads to multiple lipid transport defects. The protein may have a function in the transport of multiple lipids and their homeostasis, and may play a critical role in regulating lipid metabolism.  
CC  
CC -! SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation of brush border membranes from proximal enterocytes suggests considerable association with the apical membrane fraction. Exists as a predominantly cell surface membrane expressed protein (By similarity). According to Ref.3 localizes in a subcellular vesicular compartment rich in RAB5.  
CC  
CC -! ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=Q9UHC9-1; Sequence=Displayed;  
CC Name=2; Synonyms=NPC1L1DELTAE15;  
CC IsoId=Q9UHC9-2; Sequence=VSP\_015314;  
CC Name=3; Synonyms=NPC1LTf;  
CC IsoId=Q9UHC9-3; Sequence=VSP\_015312, VSP\_015313;  
CC -! TISSUE SPECIFICITY: Widely expressed. Expressed in liver. Also expressed in small intestine, pancreas, kidney, lung, pancreas, spleen, heart, gall bladder, brain, testis, stomach and muscle.  
CC  
CC -! INDUCTION: Expression is decreased in Caco-2 cells upon PPARdelta activation.  
CC  
CC -! PTM: Highly glycosylated (By similarity).  
CC  
CC -! POLYMORPHISM: Variations in NPC1L1 gene could be associated with nonresponse to ezetimibe treatment.  
CC  
CC -! SIMILARITY: Belongs to the patched family.  
CC  
CC -! SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

---

DR	EMBL;	AF192522;	AAF20396.1;	-;	mRNA.
DR	EMBL;	AF192523;	AAF20397.1;	-;	mRNA.
DR	EMBL;	AY152256;	AAS6939.1;	-;	mRNA.
DR	EMBL;	AY437865;	AAR97886.1;	-;	mRNA.
DR	Ensembl;	ENSG00000015520;	Homo sapiens.		
DR	HGNC;	HGNC:7898;	NPC1L1.		
DR	MIM;	608010;	-;		
DR	GO;	GO:0015918;	P:sterol transport;	TAS.	
DR	InterPro;	IPR004765;	NP_C type.		
DR	InterPro;	IPR003392;	Patched.		
DR	InterPro;	IPR000731;	SSD 5TM.		
DR	Pfam;	PF02460;	Patched; 1.		
DR	TIGRFAMS;	TIGR00917;	2A060601;	1.	
DR	PROSITE;	PS50156;	SSD; 1.		
DR	Alternative splicing;	Cholesterol metabolism;	Glycoprotein;		
KW	Lipid metabolism;	Polymorphism;	Signal;	Steroid metabolism;	
KW	Transmembrane.				
KW	SIGNAL	1	21	Potential.	
FT	CHAIN	22	1359	Niemann-Pick C1-like protein 1.	
FT	TOPO_DOM	22	284	Extracellular (Potential).	
FT	TRANSMEM	285	305	1 (Potential).	
FT	TOPO_DOM	306	351	Cytoplasmic (Potential).	
FT	TRANSMEM	352	372	2 (Potential).	
FT	TOPO_DOM	373	632	Extracellular (Potential).	
FT	TRANSMEM	633	653	3 (Potential).	

Db 301 LLVGRVAPARDKSKWDPKKGYSLSDKSFSFTHLLGGQFFQGWGTWASWPTILVLSV 360  
QY 361 IPVVALAAGLVFTTELTPDPVELWSAPNSQARSEKAFHQHFGFFRTNQVILTPAPRSSY 420  
Db 361 IPVVALAAGLVFTTELTPDPVELWSAPNSQARSEKAFHQHFGFFRTNQVILTPAPRSSY 420  
QY 421 RYDSLLGPKNFSGIILDLDDLLLELLELQERLRHLQWSPQAQRNLSIQDICVAPLNPDMT 480  
Db 421 RYDSLLGPKNFSGIILDLDDLLLELLELQERLRHLQWSPQAQRNLSIQDICVAPLNPDMT 480  
QY 481 SLYDCCINSILQYFQNNRTLLLTANTQTLWGQTSQVDWKDHFYCANAPLTPKDGTFALAL 540  
Db 481 SLYDCCINSILQYFQNNRTLLLTANTQTLWGQTSQVDWKDHFYCANAPLTPKDGTFALAL 540  
QY 541 SCWADYGAPVFPPLAIGGYKGKDYSEAEALIMTFSLNYPAGDPRLQAQAKLWEEAFLEEM 600  
Db 541 SCWADYGAPVFPPLAIGGYKGKDYSEAEALIMTFSLNYPAGDPRLQAQAKLWEEAFLEEM 600  
QY 601 RAFQRMAGFQVTFTAERSLEDEINRTTAEDLPIPATSYIVIFLYISLALGYSYSSWSRV 660  
Db 601 RAFQRMAGFQVTFTAERSLEDEINRTTAEDLPIPATSYIVIFLYISLALGYSYSSWSRV 660  
QY 661 MVDKATLGLGGVAVVLGAVMAAGPFSYLGIRSSILVILQVVPFLVLSYGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVVLGAVMAAGPFSYLGIRSSILVILQVVPFLVLSYGADNIFIFVLE 720  
QY 721 YQRLPRRPEPEVHIGRALGRVAPSMILCSSEALCFPLGALTMPAVRTTALTSGLAV 780  
Db 721 YQRLPRRPEPEVHIGRALGRVAPSMILCSSEALCFPLGALTMPAVRTTALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDKRGEASRLDVCCVKPQELPPQGGELLLGFFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDKRGEASRLDVCCVKPQELPPQGGELLLGFFQKAYAPFL 840  
QY 841 LHWITRGVVLLFLALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFVGPAPV 900  
Db 841 LHWITRGVVLLFLALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFVGPAPV 900  
QY 901 YFVTTILGVNFSSEAGNNAICSSAGCWNFSFTQKIQVATEFPQSQYLAIIPASSWVDFFIDW 960  
Db 901 YFVTTILGVNFSSEAGNNAICSSAGCWNFSFTQKIQVATEFPQSQYLAIIPASSWVDFFIDW 960  
QY 961 LTPSSCCRLYISGPNKDKPCPSTVNSLNCNCKMSITMGSVRPSVEQFHKLWFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKPCPSTVNSLNCNCKMSITMGSVRPSVEQFHKLWFLNDRP 1020  
QY 1021 NIKCPKGLAAYSTSVNLTSQGVLT-----ASRFMAYH 1053  
Db 1021 NIKCPKGLAAYSTSVNLTSQGVLT-----ASRFMAYH 1053  
QY 1054 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYITITNVFVEQYLTILPEGL 1113  
Db 1054 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYITITNVFVEQYLTILPEGL 1113  
QY 1114 FMLSCLVPTFAVSCLLGLDLRSGLNLLSVIMILVDTGPMALWDISYNAVSILNLVS 1173  
Db 1141 FMLSCLVPTFAVSCLLGLDLRSGLNLLSVIMILVDTGPMALWDISYNAVSILNLVS 1200  
QY 1174 AVGMSEVFVSHITRFAISTKTPTWLERAKEATISMGSAVFAVAMTNLPGLILVLGLAKAQ 1233  
Db 1201 AVGMSEVFVSHITRFAISTKTPTWLERAKEATISMGSAVFAVAMTNLPGLILVLGLAKAQ 1260  
QY 1234 LIQIFFRNLNLLITLGLLHGLVFLPVILSVYGGPDVNPALALEOKRAEEAAMVAVSCP 1293  
Db 1261 LIQIFFRNLNLLITLGLLHGLVFLPVILSVYGGPDVNPALALEOKRAEEAAMVAVSCP 1320  
QY 1294 NHPRSVSTADNIYVNHSEPGSKGAGAI SNFLPNNGRQF 1332  
Db 1321 NHPRSVSTADNIYVNHSEPGSKGAGAI SNFLPNNGRQF 1359

RESULT 2

NPCL1 RAT  
ID NPCL1 RAT STANDARD; PRT; 1331 AA.  
AC O6T3U3;  
DT 13-SEP-2005 (Rel. 48, Created)  
DT 13-SEP-2005 (Rel. 48, Last sequence update)  
DT 13-SEP-2005 (Rel. 48, Last annotation update)  
DE Niemann-Pick C1-like protein 1 precursor.  
GN Name=Npc1l1;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], TISSUE SPECIFICITY, AND SUBCELLULAR  
RP LOCATION.  
RC STRAIN=Sprague-Dawley;  
RX PubMed=14976338; DOI=10.1126/science.1093131;  
RA Altman S.W., Davis H.R. Jr., Zhu L.-J., Yao X., Hoos L.M.,  
Tetzloff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,  
Murgolo N., Graziano M.P.;  
RT "Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol  
absorption.";  
RL Science 303:1201-1204(2004).  
RN [2]  
RP SUBCELLULAR LOCATION, AND CHARACTERIZATION.  
RX PubMed=15777641; DOI=10.1016/j.bbagen.2004.12.021;  
RA Iyer S.P.N., Yao X., Crona J.H., Hoos L.M., Tetzloff G.,  
Davis H.R. Jr., Graziano M.P., Altman S.W.;  
RT "Characterization of the putative native and recombinant rat sterol  
transporter Niemann-Pick C1 Like 1 (NPCL1) protein.";  
RL Biochim. Biophys. Acta 1722:282-292(2005).  
RN [3]  
RP FUNCTION.  
RX PubMed=15928087; DOI=10.1073/pnas.0500269102;  
RA Garcia-Calvo M., Lisnock J., Bull H.G., Hawn B.E., Burnett D.A.,  
Braun M.P., Crona J.H., Davis H.R. Jr., Dean D.C., Demers P.A.,  
Graziano M.P., Hughes M., MacIntyre D.E., Ogawa A., O'Neill K.A.,  
Iyer S.P.N., Shevell D.E., Smith M.M., Tang Y.S., Makarewicz A.M.,  
Ujjainwalla F., Altman S.W., Chapman K.T., Thornberry N.A.;  
RT "The target of ezetimibe is Niemann-Pick C1-Like 1 (NPCL1).";  
RL Proc. Natl. Acad. Sci. U.S.A. 102:8132-8137(2005).  
CC -!- FUNCTION: Play a major role in cholesterol homeostasis. Is  
critical for the uptake of cholesterol across the plasma membrane  
of the intestinal enterocyte. Is the direct molecular target of  
ezetimibe, a drug that inhibits cholesterol absorption (By  
similarity). Lack of activity leads to multiple lipid transport  
defects. The protein may have a function in the transport of  
multiple lipids and their homeostasis.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation  
of brush border membranes from proximal enterocytes suggests  
considerable association with the apical membrane fraction. Exists  
as a predominantly cell surface membrane expressed protein.  
CC -!- TISSUE SPECIFICITY: Small intestine showed the highest level of  
expression. Expression in other tissue including gall bladder,  
liver, testis, and stomach is also observed. Along the duodenum-  
ileum axis, the levels vary in different segments of intestine,  
with peak expression in the proximal jejunum. Protein expression  
is confined to the enterocyte. Discrete localization to the  
epithelial layer bordering the luminal space along the crypt-  
villus axis. Protein expression in the enterocyte is observed  
closest to the luminal space. Expression in enterocytes from the  
proximal (jejunum) but not in the distal (ileum) region.  
CC -!- PTM: Highly glycosylated.  
CC -!- SIMILARITY: Belongs to the patched family.  
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
CC  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use as long as its content is in no way modified and this statement is not  
removed.

Mon Apr 10 07:16:16 2006

DR	EMBL; AY437867; AAR97888.1; -, mRNA.	Db	120	FVSLHCHNTCSPDQSLFINVTRVVERGAGEPPAVVAYEAFYQRFPAEYSCSRRIPA	179
DR	RGD; 1303135; Npcall1.	Qy	180	AATLAVGTMCVGYGSAALCNAORWLNFGDGTGNGLAPLDITPHLEPPGQAVSGIQLNEG	239
DR	InterPro; IPR004765; NP_C type.	Db	180	AASLAVGSMCGVYGSALCNAORWLNFGDGTGNGLAPLDITPHLEPPGQAVSGIQLNEG	239
DR	InterPro; IPR003392; PaTched.	Qy	240	VARCNESQDDVATCSQDCAASCPAIARPOALDSTFYLGQMPGSLVLIILCSFVAVVT	299
DR	InterPro; IPR000731; SSD_STM.	Db	240	IAPCNESQDDSAVCSQDCAASCFCVIPPPEALRPSFYMGWMPGMLIIITFAVFLIS	299
DR	Pfam; PF02460; Patched; 1.	Qy	300	ILLVGRVAPARDKSMVDPKGTSLSDKLSBSTHTLIGQFQGHGTWVASPLTLVLIS	359
DR	TIGRFAMs; TIGR00917; 2A060601; 1.	Db	300	AVLVRVLRVNRKNKAEGPOEAPLPHKHKLSPTILGRFFQNNGTTRVASWPLTVLIS	359
DR	PROSITE; PS0156; SSD; 1.	Qy	360	VTPVVALAAGLVFTETLTPVELWSAPNSQARSEKAFHDQHFPGPFRTNQVLTAPNRSS	419
DR	Cholesterol metabolism; Glycoprotein; Lipid metabolism; Signal;	Db	360	FIWIALAAGLTFIELTTPVLSWAPKSQARKEKSFHDEHGFPPFRTNQIFVTAENRSS	419
DR	Steroid metabolism; Transmembrane.	Qy	420	YRYDSLILGPKNFGSGLDLDLLELELQERLHRHQVMSPEAQRNISLODICYAPLNPN	479
DR	SIGNAL	Db	420	YKYSLSLLGSKNFGSGLSLDLELLELQERLHRHQVMSPEAQRNISLODICYAPLNPN	479
DR	CHAIN	Qy	480	TSLYDCINSLLOFONNRTLLLTANQTLNGQTSQVDMKDHLYCANAPLTKQGTALA	539
DR	TOPO_DOM	Db	480	TSLSDCVNSLLQYFQNNRTLLLTANQTLNGQTSQVDMKDHLYCANAPLTKQGTALA	539
DR	TRANSMEM	Qy	540	LSCMADYGAVPVPFLAIGVKGKDYSEAEALIMTFSINNYPAGDPRLAQAKLWEEAFLEE	599
DR	TOPO_DOM	Db	540	LSCMADYGAVPVPFLAIGVKGKDYSEAEALIMTFSINNYPAGDPRLAQAKLWEEAFLEE	599
DR	TRANSMEM	Qy	600	MRAFORRMAGMFQVTFATERSLEDEINRTAEDLPATPSYIVIFYIYISLALGSYSWSR	659
DR	TOPO_DOM	Db	600	MESFQNTSDKQVAFSAERSLEDEINRTIQDLPVFAVSIIYFVLYISLALGSYSRCSR	659
DR	TRANSMEM	Qy	660	VMVDSKATILGQVAVVLGVMAAMGFFSVLGRSSVLIVQVPELVLSVGADNIFIVL	719
DR	TOPO_DOM	Db	660	VAVESKATILGQVAVVLGVMAAMGFFSVLGRSSVLIVQVPELVLSVGADNIFIVL	719
DR	TRANSMEM	Qy	720	EYQRLPRRCEPEVHIGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLA	779
DR	TOPO_DOM	Db	720	EYQRLPRRCEPEVHIGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSGLA	779
DR	TRANSMEM	Qy	780	VILDFLQMGAFVALLSLDSKQEASRLDVCCCVKPOELPPQGGEGILLGFQKAYPF	839
DR	TOPO_DOM	Db	780	IILDFLQMGAFVALLSLDSKQEASRLDVCCCVKPOELPPQGGEGILLGFQKAYPF	839
DR	TRANSMEM	Qy	840	LLHWITRGVLLFLALFGVSLYSWCHISVGLDQELALPKDSYLDLYFLNRYFEYGAP	899
DR	TOPO_DOM	Db	840	LLHRTIRPVVWLLFLTLFGANLYLMCNINVLGDLQELALPKDSYLDLYFLNRYFEVGPP	899
DR	TRANSMEM	Qy	900	VYFVTTILGYNFSSAGNNAICSSAGCNFFSTQKIYATEPPEQSYLAIIPASSWDDFID	959
DR	TOPO_DOM	Db	900	VYFVTTILGYNFSSAGNNAICSSAGCNFFSTQKIYATEPPEQSYLAIIPASSWDDFID	959
DR	TRANSMEM	Qy	960	WLTP-SSCCRLYISGPNKDXPCSTVNSLNCNKMSTMGSVPSVPSVEQFKYLPWLAND	1018
DR	TOPO_DOM	Db	960	WLTPSSCCRLYIRGPHKDEPCSTDTSFNCKNKNRNLTPGVPRPTAEQFKYLPWLAND	1019
DR	TRANSMEM	Qy	1019	RPNKCPKGGAAAYSTSNLTSQGVLASRFMAWYHKLKNSQDYTEALRAARELAANITA	1078
DR	TOPO_DOM	Db	1020	PENIRCPKGGAAAYSTSNLTSQGVLASRFMAWYHKLKNSQDYTEALRAARELAANITA	1079
DR	TRANSMEM	Qy	1079	DLRKVPDTPAPEVFPFYITNVFVEQYLITLPEGLFMLSCLVPTFAVSCILLGLDLRSG	1138
DR	TOPO_DOM	Db	1080	DLRKVPDTPAPEVFPFYITNVFVEQYLITLPEGLFMLSCLVPTFAVSCILLGLDLRSG	1139
DR	TRANSMEM	Qy	1139	LINLSIYMLIVDTVGFVWALWDISYNAVSLINLSAVGMSVEFVSHITRESFAISPKPWL	1198
DR	TOPO_DOM	Db	1140	LINLSIYMLIVDTVGFVWALWDISYNAVSLINLSAVGMSVEFVSHITRESFAISPKPWL	1199
DR	TRANSMEM	Qy	1199	ERAKATISMSGSAVAGVAMTNLPGILVGLAKAQLOIIFPRLNLLTLGLLHGLVFL	1258
DR	TOPO_DOM	Db	1200	ERAKATISMSGSAVAGVAMTNLPGILVGLAKAQLOIIFPRLNLLTLGLLHGLVFL	1259

Query Match	78.5%;	Score	5421.5;	DB 1;	Length	1331;
Best Local Similarity	77.8%;	Pred. No.	0;			
Matches 1039;	Conservative	120;	Mismatches	169;	Indels	7;
Gaps	6;					
Qy	1	MAFAGLRGMLLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGLMTLSNVSCLSN	60			
Db	1	MAAALW-LGMLLWALLLSAAQGLYTPKHEGVCFTFYEECGKNPELSGLTSLSNVSCLSN	59			
Qy	61	TPARKITGDHLLILQKICPLRYTGPNTQ-ACCSAKOLVLSLEASLTKALLTRCPACSDN	119			
Db	60	TPARHVTGEHLLQKICPLRYNGPNTTFACCSKTQLLSLESSMSITKALLTRCPACSDN	119			
Qy	120	FVNLHCHNTCSPNQSLFINVTRVVAQOLGAGOLPAVAVAYEAFYQHSFAEQSDSCSRVRVPA	179			

Qy 1259 PVILSYGVDPVNPALAEQKRAEBAVAVMVASCPNHPSPRVSTADNIYVNHSPFGSI-KG 1317  
 Db 1260 PVILSYLGPVQVQALVQEKLESEA-AVAPEPSCQVQPSADADAN--VNYGFAPELAHG 1316  
 Qy 1318 AGAISNPLPNNNGROF 1332  
 Db 1317 ANAARSLPFSKQDKF 1331

RESULT 3

NPCL1\_MOUSE  
 ID NPCL1\_MOUSE STANDARD; PRT; 1333 AA.  
 AC Q6T3U4; Q5SVV1;  
 DT 13-SEP-2005 (Rel. 48, Created)  
 DT 13-SEP-2005 (Rel. 48, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Niemann-Pick C1-like protein 1 precursor.  
 GN Name=Npc1l1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muroidae; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN=C57BL/6;  
 RX PubMed=14976318; DOI=10.1126/science.1093131;  
 RA Altman S.W., Davis H.R. Jr., Zhu L.-J., Yao X., Hoos L.M.,  
 RA Tetloff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,  
 RA Murgolo N., Graziano M.P.;  
 RA "Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol  
 RT absorption.";  
 RL Science 303:1201-1204(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RG The mouse genome sequencing consortium;  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP FUNCTION, AND INDUCTION.  
 RX PubMed=15173162; DOI=10.1074/jbc.M405817200;  
 RA Davis H.R. Jr., Zhu L.-J., Hoos L.M., Tetloff G., Maguire M., Liu J.,  
 RA Yao X., Iyer S.P.N., Lam M.-H., Lund E.G., Delmers P.A.,  
 RA Graziano M.P., Altman S.W.;  
 RA "Niemann-Pick C1 Like 1 (NPCL1) is the intestinal phytosterol and  
 RT cholesterol transporter and a key modulator of whole-body cholesterol  
 RT homeostasis.";  
 RL J. Biol. Chem. 279:33586-33592(2004).  
 RN [4]  
 RP TISSUE SPECIFICITY, AND FUNCTION.  
 RX PubMed=15671032; DOI=10.1074/jbc.M409110200;  
 RA Davies J.P., Scott C., Oishi K., Liapis A., Ioannou Y.A.;  
 RA "Inactivation of NPCL1 causes multiple lipid transport defects and  
 RT protects against diet-induced hypercholesterolemia.";  
 RL J. Biol. Chem. 280:12710-12720(2005).  
 RN [5]  
 RP INDUCTION.  
 RX PubMed=15604518; DOI=10.1194/jlr.M400400-JLR200;  
 RA van der Veen J.N., Kruit J.K., Havinga R., Baller J.F.W., Chimini G.,  
 RA Lestavel S., Scaels B., Groot P.H.E., Groen A.K., Kuipers F.;  
 RA "Reduced cholesterol absorption upon PPARdelta activation coincides  
 RT with decreased intestinal expression of NPCL1.";  
 RL J. Lipid Res. 46:526-534(2005).  
 CC -1- FUNCTION: Play a major role in cholesterol homeostasis. Is  
 CC critical for the uptake of both phytosterol and cholesterol across  
 CC the plasma membrane of the intestinal enterocyte. Is the direct  
 CC molecular target of ezetimibe, a drug that inhibits cholesterol  
 CC absorption (By similarity). Lack of activity leads to multiple  
 CC lipid transport defects and exerts a protective effect against  
 CC diet-induced hyperlipidemia. Lack of activity also causes a  
 CC deregulation of CAV1 transport and localization, suggesting that  
 CC the observed lipid transport defect may be the indirect result of  
 CC an inability of cells to properly target and/or regulate CAV1

expression. The protein may have a function in the transport of  
 multiple lipids and their homeostasis, and may play a critical  
 role in regulating lipid metabolism.  
 -1- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation  
 of brush border membranes from proximal enterocytes suggests  
 considerable association with the apical membrane fraction. Exists  
 as a predominantly cell surface membrane expressed protein.  
 -1- TISSUE SPECIFICITY: Expressed in small intestine, stomach and  
 muscle, along with detectable expression in lung, heart, gall  
 bladder, brain, testis, skin and liver. Expression in liver is  
 extremely low.  
 -1- INDUCTION: Cholesterol/choleate feeding resulted in down-regulation  
 of intestinal expression. Expression is decreased by 35% in the  
 jejunum upon PPARD activation.  
 -1- PTM: Highly glycosylated (By similarity).  
 -1- SIMILARITY: Belongs to the patched family.  
 -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
 -----  
 This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 -----  
 EMBL; AY437866; AAR97887.1; -; mRNA.  
 DR EMBL; AL607152; CAI24395.1; -; Genomic\_DNA.  
 MGI; MGI:2685089; Npc1l1.  
 GO; GO:0042632; P:cholesterol homeostasis; IMP.  
 InterPro; IPR004765; NP\_Ctype.  
 InterPro; IPR003392; Patched.  
 DR InterPro; IPR000731; SSD\_5TM.  
 DR Pfam; PF02460; Patched; 1.  
 DR TIGRFAMs; TIGR00917; 2A060601; 1.  
 DR PROSITE; PS50156; SSD; 1.  
 DR Cholesterol metabolism; Glycoprotein; Lipid metabolism; Signal;  
 KW Steroid metabolism; Transmembrane.  
 KW SIGNAL 1 20 Potential.  
 FT CHAIN 21 1333 Niemann-Pick C1-like protein 1.  
 FT TOPO\_DOM 21 284 Extracellular (Potential).  
 FT TRANSMEM 285 305 1 (Potential).  
 FT TOPO\_DOM 306 352 Cytoplasmic (Potential).  
 FT TRANSMEM 353 373 2 (Potential).  
 FT TOPO\_DOM 374 632 Extracellular (Potential).  
 FT TRANSMEM 633 653 3 (Potential).  
 FT TOPO\_DOM 654 665 Cytoplasmic (Potential).  
 FT TRANSMEM 666 686 4 (Potential).  
 FT TOPO\_DOM 687 717 Extracellular (Potential).  
 FT TRANSMEM 718 742 5 (Potential).  
 FT TOPO\_DOM 743 763 Cytoplasmic (Potential).  
 FT TRANSMEM 764 776 Extracellular (Potential).  
 FT TOPO\_DOM 777 797 7 (Potential).  
 FT TRANSMEM 798 846 Cytoplasmic (Potential).  
 FT TOPO\_DOM 847 867 8 (Potential).  
 FT TRANSMEM 868 1113 Extracellular (Potential).  
 FT TOPO\_DOM 1114 1134 9 (Potential).  
 FT TRANSMEM 1135 1142 Cytoplasmic (Potential).  
 FT TOPO\_DOM 1143 1163 10 (Potential).  
 FT TRANSMEM 1164 1165 Extracellular (Potential).  
 FT TOPO\_DOM 1166 1186 11 (Potential).  
 FT TRANSMEM 1187 1206 Cytoplasmic (Potential).  
 FT TOPO\_DOM 1207 1227 12 (Potential).  
 FT TRANSMEM 1228 1242 Extracellular (Potential).  
 FT TOPO\_DOM 1243 1263 13 (Potential).  
 FT TRANSMEM 1264 1333 Cytoplasmic (Potential).  
 FT DOMAIN 632 797 SSD.  
 FT COMEBIAS 267 270 Poly-Leu.  
 FT COMEBIAS 500 503 Poly-Leu.  
 FT CONFLICT 476 476 N -> K (in Ref. 2).  
 SQ SEQUENCE 1333 AA; 147132 MW; 77715209B352735 CRC64;

Query Match 78.3%; Score 5407; DB 1; Length 1333;  
 Best Local Similarity 77.2%; Pred. No. 0;

Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;			
Qy	4	AGLRGWLWALLRLAQSEPTTTHQPGYCAFYDECKNPKLSGLMTLSNVCSLNTPA	63
Db	3	AAWQGLLWALLRLAQSEPTTTHQKAGCTTFEECCGNPKLSGLTSLNISCLNTPA	62
Qy	64	RKITGDHLLILOKTCPLRYTGNP-TOACCSAKQLVSLASLSITKALLTRCPACSDNFVN	122
Db	63	RHVITGDHLLALQRCVPRLYNGPNDTYACCSKQLVSLDSSLISITKALLTRCPACSENFVS	122
Qy	123	LHCNHTCSPNQSFLNTRVQAQAGQLPAVVAEYQHSFAQSCVDSRVRVPAAT	182
Db	123	IHCNHTCSPQSLFNTRVVRQDQGLPAVVAEYQHSFAQSCVDSRVRVPAAS	182
Qy	193	LAVGTMCGVYSGALCNAQRWLNFGDGTGNGLAPDITPHLEPGQAVGSGIQLNEGVAR	242
Db	193	LAVGTMCGVYSGALCNAQRWLNFGDGTGNGLAPDITPHLEPGQADGKPLDGKITP	242
Qy	243	CNESQDDVATCSCQDCAASCFAIARPOALDSTFYLGMPGSLVLIILICSVFVAVTILL	302
Db	243	CNESQEDSAASCQDCAASCPIVPPALRPSFYMGMRPGWLAIIIFTAVFVLLSVVL	302
Qy	303	VGRVAPARDKSMVDPKGTSLSKLSFSHTLLGQFQCGWTVASWPLTILVLSVIP	362
Db	303	VYLRVASNRKNKTAGSQEAPNLPKRFRSPHTVLGRFFESWGTQVASWPLITVLSFIV	362
Qy	363	VVALAAGLVFTETLTPDVELWAPNSQARSEKAFHQHFGPPFRTNQVILTPAPNRSYRY	422
Db	363	VIALSVGLTFTIETLTPDVELWAPKSAQKAKAFHDEHFGPPFRTNQIVFTAKNRSYKY	422
Qy	423	DSILLGPKNFSGILDLLELLELQERLRLHQLVQWSPRQARNISLQDIQYAPLNPDTSL	482
Db	423	DSILLGPKNFSGILSLDLLELLELQERLRLHQLVQWSPRQARNISLQDIQYAPLNPHTSL	482
Qy	483	YDCCINSLLQYQONNRTLLLTANQTLMGOTSQVDWKHFLCANAPLTFKDGKTALASC	542
Db	483	TDCCVNSLLQYQONNHTLLLTANQTLNGOTSLVDWKHFLCANAPLTYKDGKTALASC	542
Qy	543	MADYGAPVFPFLAIGYKGDYSEAEALIMTSLNNYPAGDPRLAQAKLWESAFLEWRA	602
Db	543	IADYGAPVFPFLAVGYQGTDYSEAEALITFSINNPADDPMAHAKLWESAFLEWQMS	602
Qy	603	FORMAGMFQVPTAERSLEDEINRTAEDLPFATSIYIVILYISLAGSYSSRSRYMV	662
Db	603	FORSTADKQIAFSAERSLEDEINRTTIQDLPVFAISYILYISLAGSYSSRSRYMV	662
Qy	663	DSKATIGLGVAVVILGAVMAAGFFSVGLIRSSLVILQVVPFLVLSVGADNIFIFVLEYQ	722
Db	663	DSKATIGLGVAVVILGAVMAAGFFSVGLIRSSLVILQVVPFLVLSVGADNIFIFVLEYQ	722
Qy	723	RLPRRPGEPREVHIGRALGRVAFPSMLLCSLSEAICFFLGALTTPMPAVRTFALTSLGAVIL	782
Db	723	RLPRMPEQREAHIGRTLGSVAFPSMLLCSLSEAICFFLGALTSPMPAVRTFALTSLGAIIF	782
Qy	783	DFLLQMSAFVALLSLDSKROEASRLDVCCCKQELPPQCGGSLILGPFQKAYAPFLH	842
Db	783	DFLLQMTAFVALLSLDSKROEASRPDVCCFSRRNUPPPKQKGLLCCFFPKLYTFPLH	842
Qy	843	WITGVVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFFLNFYFVGVAPVYF	902
Db	843	RFIRPVVLLLFVLFGANLYLMCNISVGLDQDLALPKDSYLLDYFFLNFYFVGVPPVYF	902
Qy	903	VTTLYGNFSSBAGMNAICSSAGCNFNFSFTQKIQVATFEPQSYLAIPASSWVDFTDMLT	962
Db	903	DTTSGYNFSTEAGMNAICSSAGCESFSLTQKIQVASEFPNQSYVIAAASSWVDFTDMLT	962
Qy	963	P-SSCCRIYISGPNKDECPSTVNSLCKNCSITWGSVRPSVEQFKHLYPFLNDRPN	1021
Db	963	PSSSCRIYTRGPHKDECPSTDTSFNCKNCKMRTLGPVPTTEQFKHLYPFLNDRPN	1022
Qy	1022	IKCPKGLLAAYSTSVNLTSQVILASRFMAVYHFKLNSQDYTEALRAARELAANITADLR	1081
Db	1023	IRCPKGLLAAYSTSVNLSSDQIIIASQFMVYHFKLNSQDYTEALRASRLIAANITAE LR	1082

Qy	1082	KVPGTDPAFEPVPTITNVFEYOYLITLPEGLFMLSCLVPTFAVSCILGLDIRSGILN	1141
Db	1083	KVPGTDPAFEPVPTITNSVYQOYLTVLPEGITLALCFVPTFVVCYLLGLDIRSGILN	1142
Qy	1142	LLSVIMILVDITVGFMAIWDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLERA	1201
Db	1143	LLSVIMILVDITIGLMAVVGISYNAVSLINLVTAVGMSVEFVSHITRSFAVSTKPTLERA	1202
Qy	1202	KEATISMGSAVFAVAVMTNLPGLIVLGLAKAQITQIFFERLNLITLLGLLHGLVFLPVI	1261
Db	1203	KDQATIFMGSAVFAVAVMTNFPGLITLGLFAQAQIQLIFFERLNLITLLGLLHGLVFLPVV	1262
Qy	1262	LSYGVDPNVPALALEOKRAEEAAVAAVMVASCNHPHSRSTADNTYVNHSPGS-IKGAGA	1320
Db	1263	LSYLGDPVNVQALVLEEKATEA-AMVSEPSQCYPPFPADANTSDYVNYGFNPEFPEINA	1321
Qy	1321	ISNLPNNGRQF 1332	
Db	1322	ASSSLPKSQDKF 1333	
RESULT 4			
ID	Q4T749	TETNG	
AC	Q4T749	TETNG PRELIMINARY; PRT; 1132 AA.	
DT	13-SEP-2005	(T-EMBLrel. 31, Created)	
DT	13-SEP-2005	(T-EMBLrel. 31, Last sequence update)	
DT	13-SEP-2005	(T-EMBLrel. 31, Last annotation update)	
DE	Chromosome undetermined SCAF8317, whole genome shotgun sequence.		
DE	(Fragment)		
OS	ORFNames=GSTENG0005930001;		
GN	Tetraodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontoidea; Tetraodontidae; Tetraodon.		
OX	NCBI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jallion O., Aury J.M., Brunet P., Petit J.L., Stange-Thomann N.,		
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Coestaz C., Bernot A.,		
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,		
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,		
RA	Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,		
RA	Biemont C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,		
RA	Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,		
RA	Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,		
RA	Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpetti C.,		
RA	Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;		
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals		
RT	the early vertebrate proto-karyotype.";		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope; Whitehead Institute Centre for Genome Research;		
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.		
CC	-!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CAAB01008317; CAF91283.1; -; Genomic_DNA.		
FT	NON TER 1		
SQ	SEQUENCE 1132 AA; 125923 MW; 4283CF4CA0BBDE97 CRC64;		
Query Match 46.0%; Score 3179; DB 2; Length 1132;			
Best Local Similarity 52.0%; Pred. No. 4.3e-216;			
Matches 641; Conservative 167; Mismatches 264; Indels 160; Gaps 11;			
Qy	87	TOACCSAKQLVSLASLSITKALLTRCPACSDNFVNHLCHNTCSNPQSLFINTRVQAQL-	145
Db	1	TYACCSINQLSSLEASLASKAVLIRCPSCADNFAHFHCITTCSPQQTIKVTKVNTV	60

Qy	146	GAGQL--PAVWAYEYQHSFAEQSVDSCSRVRVPAATLAVTGWGVYGSALCNAORWL	203
Db	61	GFDNITRGVGVYAYLSTNFADASFQSCKNVRIIPATGGLATIGTWGRPGSRHCNPQRY	120
Qy	204	NFQDGTGNGLAPLDTTFHLLBERG--QAVGSGIQPLNEGVARCNESQGDQVATCSQDCAA	261
Db	121	DFQGDSSNGLAPLDDFHLIRKGDTEDELPBGIVPYVSGNALKCNETPTSGGRKACSQDQOE	180
Qy	262	SCPATARQALDSTFYLCQMPGSLVL--IIILC-SVPAVVTTLLVGF-----RV	307
Db	181	SCPKITPPPLPPGPPQLLGMDFLLIISIILCLLIFAFELFVFLVTVYLRNNNNKGLKKKD	240
Qy	308	APARDKSNVDPKKGTSLSDKLSPFSTHTLLGQFFQMGWTWASWPLTILVLSVIPVVALA	367
Db	241	QKSHDVORVISPSEVTCADKNLSQAQAPLSKSFQIWGTWATYPLTWLLLSAIIIVVWLS	300
Qy	368	AGLVFTELTTDPVELWSAPNSQARSEKAFHQHFGFFERTNQVILITAPNRSYRVDSLIL	427
Db	301	AGIKNIELTTDPVELWSAPNSRARQEKAFHDTHPDPTFNQILITAPKSGYMWDSLIF	360
Qy	428	GPKNPSGILDLLLELLELQERLHLQWSPQAQRNLSLODICVAPLNPNONTSLYDCCI	487
Db	361	GOHNLSGIISKIDLLIELELOKQINIEFWSKDNLRATLKDVCFAPLNPNASLTDCAV	420
Qy	488	NSLLQYFONNRTLLLTANQTLMGQTSOVDKDHFLYCANAPLTFKDTALALSCMADYG	547
Db	421	NSLPQYFONSLENIINAKVMNTELGVTKEVDNRDHLIYCFNSPLSPKDTIDLGMSCMADYG	480
Qy	548	APVPPFLAIGYKGKDYSEABALIMTFLSNVYPAGDPLAQAOKLWEAEFLBEMRAFORM	607
Db	481	APVIPFLAVGGYQ-----NAMEWEAEFLKIVQEQKNP	513
Qy	608	AGMPOVTTAERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSYSSWGRM-VDSKA	666
Db	514	STNFTLAYMAERSLEDEINRTAEDIPIMISYAVIFLYIAVALGEFTSWKRILVVDSKF	573
Qy	667	TLGLGGVAVLGAVMAGPFSYLGRSSVLQVWPPELVSVGADNIFIFVLEVORLPR	726
Db	574	LVGLGGLVSCAVLSSLGFCSWGIPPSLLIVQVVPFLVAVGADNIFIFVLEVQRDVR	633
Qy	727	RPGEPREVHIGRALGRVAPSMLLCSLSBAICFFLGALTPMPAVRTFALTGSLVILDLL	786
Db	634	RPGEKREEQIRILGNVAPSMLLCSLSSEICFFFGALSTMPAVKSPALYAVALLDPAL	693
Qy	787	QMSAFVALLSDSRQEBASRLDVCCKVPQBLPPGQGEGLLLGFFQKAYAPFLHMITR	846
Db	694	QMTAFVALLSDCRQDSNRCELLCCCIKVSRSKRTKPNEGFLLPFMKKYIYAPALINRYSR	753
Qy	847	GVVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLNRFVFCVAPYFVTTL	906
Db	754	IIVNFVTFMLCGSIFLHFHVKGLDQELAMPKDSYMLKYPEYIKYKPKVGAPYFVTKR	813
Qy	907	GYNFSSEAGMNAICSSAGCNFSTQKIQTATEPPEQSYLAIPASSWVDDFDIDLTPSSC	966
Db	814	GYNFTSVSGMNAVCSVSGCDPYSFYTKIQIATYEPQLQOLDL-----855	
Qy	967	CRLYISGNKDKFCPSTVNSLNCNLKNCMSITMGVSVRPSVEQFHKLWPFLNDRNRNIKCPK	1026
Db	856	-----IRKMIN	860
Qy	1027	GGLAAYSTSVMLTSDGQVLASRFWAYHKPLKNSODYTEALRAARELAANITADLRKVPCT	1086
Db	861	SPMLIYANDFG-----ASRFWAYHTPLTNSQFTGALMKARELAHNITMAMRQIQGT	912
Qy	1087	DPAPFVFPYTT-----ITNVFVEQVITILPBGELF	1114
Db	913	DPNEFVFPYTVVSLFSILPLSTYSCTDFTIHNDCVFSILPPARVTVNFVEQVITIVPEGLV	972
Qy	1115	MLSCLVPTFAVSCILGLDLSRGLMLNLSIIVMLIVDTVPGMALWDTSYNVAVSLINLVS	1174
Db	973	IISLCLPTFVCCLLGLDLSRGLMLNLTITIMTVDVTGWTMLWGDYNAVALINLWTA	1032

Qy	1175	VGMSEVFVSHITSPALSTKPTWLERAKENTISMGSAVPAVGAVNTMLPGILVTLGLAKAQL	12334
Db	1033	VGISVFVSWHTSRFALSISKPTTVERAKEATQMGSAVPAVGAVNTMLPGILVTLAFKAQL	1092
Qy	1235	IQIPFFPRNLNLTLLGLLHGLVFLPVLISYVG	1266
Db	1093	IQIPFFPRNLVITLLGMAHGLIFLPVLISYFG	1124
RESULT 5			
ID	Q9JLG3	CRIGR PRELIMINARY;	PRT; 1277 AA.
AC	Q9JIG3;		
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	
DT	01-OCT-2000	(TrEMBLrel. 15, Last sequence update)	
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)	
DE	Niemann-Pick type C1 protein.		
GN	Name=NPC1;		
OS	Cricetulus griseus (Chinese hamster).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;		
OC	Muridae; Cricetinae; Cricetulus.		
OX	NCBI_TaxID=10029;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	TISSUE=Ovary;		
RX	MEDLINE=20127878; PubMed=10660558; DOI=10.1074/jbc.275.6.4013;		
RA	Cruz J.C., Sugii S., Yu C., Chang T.-Y.;		
RT	"Role of Niemann-Pick type C1 protein in intracellular trafficking of		
RT	low density lipoprotein-derived cholesterol.";		
RL	J. Biol. Chem. 275:4013-4021(2000)		
DR	EWBL: AF182744; AAF31592.1; -, mRNA.		
DR	GO: GO:0016021; C:integral to membrane; IEA.		
DR	GO: GO:0005764; C:lysosome; IEA.		
DR	GO: GO:0016020; C:membrane; IEA.		
DR	GO: GO:0008158; F:hedgehog receptor activity; IEA.		
DR	GO: GO:0003031; P:cholesterol transport; IEA.		
DR	InterPro: IPR004765; NP C type.		
DR	InterPro: IPR001392; Patched.		
DR	InterPro: IPR000731; SSD_5TM.		
DR	Pfam: PF02460; Patched; I.		
DR	TIGRFAMs: TIGR00917; 2A060601; 1.		
DR	PROSITE: PS0156; SSD; 1.		
SQ	SEQUENCE 1277 AA; 142817 MW; 6B3DG38094B4E156 CRC64;		

Qy	1175	VGMSVFEVSHITSPASTSTKPTWLERAKENTISWGSAGVAGVANTMLPGTILVGLAKAQL	1234
Db	1033	VGLSVFEVSHITSPASTSTKPTWLERAKENTISWGSAGVAGVANTMLPGTILVGLAKAQL	1092
Qy	1235	IQIPFFPRLNLLITLGLLHLGLVLPVLVSYG	1266
Db	1093	IQIPFFPRLNLLITLGLLHLGLVLPVLVSYG	1124

Qy	1175	VGMSVFEVSHITSPASTSTKPTWLERAKENTISWGSAGVAGVANTMLPGTILVGLAKAQL	1234
Db	1033	VGLSVFEVSHITSPASTSTKPTWLERAKENTISWGSAGVAGVANTMLPGTILVGLAKAQL	1092
Qy	1235	IQIPFFPRLNLLITLGLLHLGLVLPVLVSYG	1266
Db	1093	IQIPFFPRLNLLITLGLLHLGLVLPVLVSYG	1124



[illegible]





```
Db      285 GAFFALWCYRKRYFVSEYTPIDSNIAFVSNANDRGE-----ASCCDALGAAPFEGCLRR 337
Qy      340 FFGQGWVWASWPLTIIVLVSIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAFHQD 399
Db      338 LFSQWGSFCVRNPGPIIFSLAFATACSSGLVFRVTNNPDLWSAPSSQARLEKEYFDT 397
Qy      400 HFGPFFRTNQVILPAPNRSSVRYDSLLGPK-NPFGSILDLDDLLELLELLELRLHLQWS 458
Db      398 HFGPFFRTEQLIQAPHTSAHTYQPYPSGSDVPFGPDLAILHQVLDLQTAIEN--ITA 455
Qy      459 PEAQNRSLQDICYAPLNPNTSLYDCINSLLQYFONNRLLALLTANQTLMGQTSOV-- 516
Db      456 SYNNETVTLQDVCVAPSPYNNK---NCTILSVLNYFQNSHSMLE-----DHEIGDDFFVYA 507
Qy      517 DWKDHFLYCANAPITFDGKTALALSCMADYGAPVFPFLAIGYKGYKDYSEAEALIMTFSL 576
Db      508 DYHTHLICYRPAASLNDTSLHDPCLGTFGFPVFWMLVGLGYDDQYNNATALVITFPV 567
Qy      577 NNYPAGDPRLAQAKLWBEAFLEENRAFQRRMAGMQVTTFAERLSLEDEINRTAEDLPIF 636
Db      568 NNYNDYTERLQKAHWEKEFTNFVKNYKN---PNLTISFTTTERSIEDELNRESNGDIFTV 624
Qy      637 ATSYIVIFLYISLALGSSSRVWDSKATLGLGGVAVLGAVMAGPSPSYLGIRSSL 696
Db      625 IISVAIMFLYISIALGHKICSRLLVDSKISLGIAGILIVLSSVACSLGIFSYVIGLPLT 684
Qy      697 VILQVWPLVLSVGADNIFIVLEYQRLPRPGPBPREVHIGRALGRVAPSMLLCSLSEAI 756
Db      685 IVIEVIFVLAVGVDNIFILVQYQDRERLHGETLQDLQGLVGEVAPSMFLSSFEAV 744
Qy      757 CFFLGALTMPAVRTFALTGLAVILDFLLQMSAFVALLSDLSKQZASRLDVCCKVQP 816
Db      745 AFFLGALSKMPAVHTSLFAGWAVLIDFLQITCFVSLGLDILKQEKRLNDVLCCVRGS 804
Qy      817 ELPPPGQ-GEGLLGFQKAYAPFLHWTGCVLLFLALFGVLSYMSCHISVGLDQEL 875
Db      805 EDGTSVQASECLFRLFKHSYSPLLLDKWMRPVIALFVGLVSFVAVLNKVEIQLDQSL 864
Qy      876 ALPKDSYLLDYFLFLNRYFREGAPVYFVTTILGYNFSSEAGNAIACSSAGCNFFSTQKI 935
Db      865 SMPDSDSYMDYFKSL-KYLHAGPPVYFVLEBGRHDYTSLKQGNMVCWGMCNDSLVQI 923
Qy      936 YATFPPEOSYLAIPASSWDDPDIWLTP-SSCRLYISGPKDKCEPSTVNSLCKNKM 994
Db      924 NAAQLDSYTRIGFAPSSWIDYDFWVKPQSSCCRYSN---TDRFCNASVVDPAICR-CR 979
Qy      995 SITM-GSVRPSVEOPHKYLPWFLNDRPNIKCPKGLAAYSVN-LTSDGOVLASRFMAY 1052
Db      980 PLTQEGKQRPQGGDFMRFLPFLSDNPNPKCGKGHAAYSAVNLLGNDTGVGATYFTY 1039
Qy      1053 HKPLKNSQDYTEALRAARELANITADLRKVPGETDPA-FEVPFYTTITNVFYEQYLTLPE 1111
Db      1040 HTVLQTSADFTDAMEKANLIASNIT---KTMGLEGSNRYFVPSVVFVEYEQYLTIID 1095
Qy      1112 GLFMLSCLVPTFAVSCLLGLDLBSGLNLLSYVMILVDVTGFMALWDISYNAVSLNL 1171
Db      1096 TIFNLVSGLGAIFLTVTVILGCDLWSAVIMCITTAMILVNNFVGMWMLGISLNAVSLNL 1155
Qy      1172 VSAGCMSVEFVSHITRSFAISTKPTWLERAKEATISMGSAFVAGVAMTNLPGIILVGLAK 1231
Db      1156 VMSCGISVEFCSHITRATTVSMKSGRAQAREALAHMGSSVFSGITLTKFGGIVLVAFAK 1215
Qy      1232 AQLIQIFFRNLMLITLLGLHLGLVFLVPLVILSYVGPDPVNPALAL---EQKRAE 1282
Db      1216 SQIFQIFFRYLANVLLGATHGLIFLPLVLLSYIGFSINKAKSLATQOYKGTG 1269
```

RESULT 8  
Q9GK52 CANFA  
ID Q9GK52 CANFA PRELIMINARY; prt; 1276 AA.  
AC Q9GK52  
DT 01-MAR-2001 (TREMREL. 16, Created)

```
DT      01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT      01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE      Niemann-Pick type C1 disease protein.  
GN      Name=NPCL;  
OS      Canis familiaris (Dog).  
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC      Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
OC      Canis.  
OX      NCBI_TaxID=9615;  
RN      [1]  
RP      NUCLEOTIDE SEQUENCE.  
RA      Murakawa M., Freeman M.W.;  
RL      Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR      EMBL; AF315034; AAC40873.1; -, mRNA.  
DR      Ensembl; ENSCAFG00000018183; Canis familiaris.  
DR      GO; GO:0016021; C:integral to membrane; IEA.  
DR      GO; GO:0005764; C:lysosome; IEA.  
DR      GO; GO:0016020; C:membrane; IEA.  
DR      GO; GO:0008158; P:hedegohg receptor activity; IEA.  
DR      GO; GO:0030301; P:cholesterol transport; IEA.  
DR      InterPro; IPR004765; NP_C-type.  
DR      InterPro; IPR003392; Patched.  
DR      InterPro; IPR000731; SSD 5TM.  
DR      TIGRFAMs; TIGR00917; 2A060601; 1.  
DR      PROSITE; PS0156; SSD; 1.  
SQ      SEQUENCE 1276 AA; 141695 MW; 07D1327B51BD5681 CRC64;  
  
Query Match 34.8%; Score 2404.5; DB 2; Length 1276;  
Best Local Similarity 40.4%; Pred. No. 4.2e+161;  
Matches 524; Conservative 233; Mismatches 434; Indels 105; Gaps 28;  
  
Qy      33 CAFYDECGKNPGLSGLMTLSNVSCLSNTPARKITGTHLLILQKICRLTYGPNPTQACS 92  
Db      25 CWVYGECC---IASGD---KRYNCQYSGPPKPLKDGVDLMQSLCPLFP-DNVSVCCD 76  
Qy      93 AKQLSVLEASLITKALLTRCPACSDNFVNLHCHTCSPNQSLFINVTR----VAQLGAG 148  
Db      77 VQQLRTKDLQQLPLQFLSRCPSCFYNNMLFCELTCSQRSQFLNTEVEDYDPTNQ 136  
Qy      149 QLPVAVAYEAFYQHSFAEQSYDSCSRVRVPAATLAVGTWCGVYGSALCNAQRWLNFGD 208  
Db      137 TKTNVKELQYVGVESFANAMYNACRDVEAPSSNDKALGLLCGKEAEA-CNATNMIYMFN 195  
Qy      209 TGNGLAPL-----DITFHLLPGQAVGSGIQPLNEGVARCNESGDDVATSCQCAA 261  
Db      196 KDNQAPFTIIPISDLPAH-----GMKPMNNAATKGCDFPDEVTPAPCSQDCSV 245  
Qy      262 SC-----PATARQALDSTFYLQMPGSLVLIILCSVFAVVVITLLVGFVAPARD 312  
Db      246 VCGPKPQPPAPAPWRILGLDAMVIMWIYMAFLMFFGAFPAWCY----- 293  
Qy      313 KSKMVDPKGTSLSDKLSFSFTH-----LLGQFFQG-----WGTWVASWPLT 354  
Db      294 -RKRYFVSEVTPIDSNIAFVSNAGDTGEASCCDAGAAFEGLRRLFTQWGSFCIRNPGC 352  
Qy      355 ILVLSVTPVVALAAGLVFTLTTPVELWSAPNSQARSEKAFDHOHGFPRPTQVILTA 414  
Db      353 IIFSLAFIAACSSGLVFSRVTTNPDLWSAPGSOARLEKEYFQFAHGFPPFRTQELIIQA 412  
Qy      415 PNRSYRYDSLILGPK-NFGSILDLDLLELLELLELLELLELLELLELLELLELLELLEL 473  
Db      413 PHTSHYVQPPSSDVPFGPPLDIGILHQVLDLQTAIEN--ITATYNNETVTLQDICA 470  
Qy      474 PLNPDTSLYDCCNSLLQYFQNNRTLLLLTANQTLMGQTSQV--DWKDHFLYCANAPLT 531  
Db      471 PLSYNNK---NCTIMSVLNYFQNSHSMLE-----DHKIGDDFVYVADYHTHLLYCVRA 522  
Qy      532 FKDGTALALSCMADYGAPVFPFLAIGYKGYKDYSEAEALIMTFSLNNYPAGDPRLAQAKL 591  
Db      523 LNDTSLHDPCLGTFGGFVFWLVLGGYDDQYNNATALVITFPVNNYNTDEKLRQAQA 582  
Qy      592 WEEAFLEEMRAFQRRMAGMQVTTFAERLSLEDEINRTAEDLPATSYIVIFLYISLAL 651
```

Db 583 WEKEFINFKNYEN---PNTISFTTRESGDELNRESGDFVFTLISYAVMFLYISIAL 639  
Qy 652 GSYSSWGRVVDKATLGLGVAVVLGAVMAAGFFSYLGRSSLVLTQVVPFLVLSVGA 711  
Db 640 GHKSCSRFLVDSKISGIAGILVLSVMCSGIFSYFGIPLTLVIEVPLFLVLAGV 699  
Qy 712 DNIPIFVLEYORLPRRGEPREVHIGRALGRVAPSMLLCSLSBAICFFLGALTMPAVRT 771  
Db 700 DNIPIFVLEYORLPRRGEPREVHIGRALGRVAPSMLLCSLSBAICFFLGALTMPAVRT 759  
Qy 772 FALTSLGAVLDELQMSAFVALLSDSKQESRLDYCCVXPQELPPPGQG-----EGL 827  
Db 760 FSLFAGVAVLDELQITCFVSLGLDLKQEKRLDLVCLCTGSE---GGTGIQASESC 816  
Qy 828 LLGFFOKAYAPFLLHMTITRGVLLFLALFVSLYSMCHTSVGLDQSLALPKDSYLLDYF 887  
Db 817 LFRFFKNSYSPFLKQWMRPIVATVFGIILSFIAVLNKVEIGLDQSLSDPDYSYMDYF 876  
Qy 888 LFLNRYEVGAPVYFVTTLGYNFSSEAGMNAICSSAGCNFNFSFTQKIYATEPPEOSYLA 947  
Db 877 KSL-KYLHAGPPVYFVLEEGHDYTSLEGQNMVCGMGNDSLVQQIFSAQLDNYTHIG 935  
Qy 948 IPASSWDDPIDMLTP-SSCRLIYISGNKDKFCSTPSTNSLNCGLKNCMSITM-GSVRPSV 1005  
Db 936 FAPSSWIDDYFDWVKPQSSCCRYNS---TDQFCNASVVDPAVR-CRPLTQEGKRRPQG 991  
Qy 1006 EQPHKYLWFLNDRPNKCPKGLAAYSTSVNLT-SDQVLSRFRWAYHKKPLANSODYTE 1064  
Db 992 EDFWFLPMLNSPNPKCKGKGAAYGSAVNLVGNDSVGTATFYHTVLTQTSADFTD 1051  
Qy 1065 ALBAARELANITADLRKRVGTDPAPEVFYTTITNVEQYLTLPEGLFMLSCLCLVPTF 1124  
Db 1052 AMRKAILIASNITKTM-SLKGSH--YRPPYSYFVYFYEQYLTLIIDTIFNLVSLGAIF 1108  
Qy 1125 AVSCLILGLDLRGLNLISIVMLIVDTVGFPMALWDISYNAVSLINLVAAGMSVFEVSH 1184  
Db 1109 LVTVLVLGCELWSAVTMCVTIAMILVNMFGVMWLWGISLNAVSLVNLVMSCGISVFEVSH 1168  
Qy 1185 ITRSPAISYKPTWLERAKETATSMGSAVAGVAMTLPGLVLGLAKAQLOIQIFPRLNL 1244  
Db 1169 ITRAFVTSAGSVEREAREALSHGSSVFGITLTFGGLVLAFAKSLQIFQVFTGMYL 1228  
Qy 1245 LITLGLGLHGLVPLVILSVGPDVNPALAL-EQKR 1279  
Db 1229 AMVLGATHGLIFLPVLLSVIGPSINKAKSLASQER 1264

## RESULT 9

ID NPC1 HUMAN  
AC O15118; Q9P130; STANDARD; PRT; 1278 AA.  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 10-MAY-2005 (Rel. 47, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Niemann-Pick C1 protein precursor.  
GN Name=NPC1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE, VARIANT ILE-642, AND VARIANTS NPC1.  
RX MEDLINE=97362323; PubMed=9211849; DOI=10.1126/science.277.5323.228;  
RA Carstea E.D., Morris J.A., Coleman K.G., Loftus S.K., Zhang D.,  
RA Cummings C., Gu J., Rosenfeld M.A., Pavan W.J., Krizman D.B.,  
RA Nagle J., Polymeropoulos M.H., Sturley S.L., Ioannou Y.A.,  
RA Higgins M.E., Comly M., Cooney A., Brown A., Kaneski C.R.,  
RA Blanchette-Mackie E.J., Dwyer N.K., Neufeld E.B., Chang T.-Y.,  
RA Liscum L., Strauss J.F. III, Ohno K., Zeigler M., Carmi R., Sokol J.,  
RA Markie D., O'Neill R.R., van Diggelen O.P., Ellender M.,  
RA Patterson M.C., Brady R.O., Vanier M.T., Pentchev P.G., Tagle D.A.;

RT "Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis." Science 277:228-231 (1997).  
RL [2]  
RN NUCLEOTIDE SEQUENCE, AND VARIANTS.  
RX MEDLINE=99355599; PubMed=10425213; DOI=10.1006/bbrc.1999.1070;  
RA Morris J.A., Zhang D., Coleman K.G., Nagle J., Pentchev P.G., Carstea E.D.;  
RA "The genomic organization and polymorphism analysis of the human Niemann-Pick C1 gene." Biochem. Biophys. Res. Commun. 261:493-498 (1999).  
RL [3]  
RN NUCLEOTIDE SEQUENCE, AND VARIANT ILE-642.  
RX MEDLINE=21623216; PubMed=11754101; DOI=10.1002/humu.10016;  
RA Bauer P., Knoblich R., Bauer C., Finckh U., Hufen A., Kropp J., Braun S., Kustermann-Kuhn B., Schmidt D., Harzer K., Rolfs A.;  
RT "NPC1: complete genomic sequence, mutation analysis, and characterization of haplotypes." Hum. Mutat. 19:30-38 (2002).  
RL [4]  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT ILE-642.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RL [5]  
RN CHARACTERIZATION.  
RX MEDLINE=99128318; PubMed=9927649; DOI=10.1073/pnas.96.3.805;  
RA Watari H., Blanchette-Mackie E.J., Dwyer N.K., Glick J.M., Patel S., Neufeld E.B., Brady R.O., Pentchev P.G., Strauss J.F. III;  
RT "Niemann-Pick C1 protein: obligatory roles for N-terminal domains and lysosomal targeting in cholesterol mobilization." Proc. Natl. Acad. Sci. U.S.A. 96:805-810 (1999).  
RL [6]  
RN VARIANT NPD TRP-992.  
RX MEDLINE=9829797; PubMed=9634529;  
RA Greer W.L., Riddell D.C., Gillan T.L., Girouard G.S., Sparrow S.M., Byers D.M., Dobson M.J., Neumann P.E.;  
RT "The Nova Scotia (type D) form of Niemann-Pick disease is caused by a G3097-->T transversion in NPC1." Am. J. Hum. Genet. 63:52-54 (1998).  
RN [7]  
RP VARIANTS NPC1 GLN-934; LEU-940; ASN-948; LEU-954; TRP-992; ALA-1007; THR-1061 AND VAL-1213.  
RX MEDLINE=99452586; PubMed=10521290;  
RA Greer W.L., Dobson M.J., Girouard G.S., Byers D.M., Riddell D.C., Neumann P.E.;  
RT "Mutations in NPC1 highlight a conserved NPC1-specific cysteine-rich domain." Am. J. Hum. Genet. 65:1252-1260 (1999).  
RN [8]  
RP VARIANT NPC1 THR-1061.  
RX MEDLINE=99452593; PubMed=10521297;  
RA Millat G., Marçais C., Rafi M.A., Yamamoto T., Morris J.A., Pentchev P.G., Ohno K., Wenger D.A., Vanier M.T.;

"Niemann-Pick C1 disease: the I1061T substitution is a frequent mutant allele in patients of Western European descent and correlates with a classic juvenile phenotype.";  
Am. J. Hum. Genet. 65:1321-1329(1999).  
[9]  
RN VARIANTS NPC1, AND VARIANTS ARG-215; VAL-858 AND GLN-1266.  
RX MEDLINE=99408226; PubMed=10480349; DOI=10.1007/s004390051057; Zhang H., Yamamoto T., Namba E., Ninomiya H., Higaki K., Taniguchi M., Zhang H., Akaboshi S., Watanabe Y., Takeshima T., Inui K., Okada S., Tanaka A., Sakuragawa N., Millat G., Vanier M.T., Morris J.A., Pentchev P.G., Ohno K.;  
RA "NPC1 gene mutations in Japanese patients with Niemann-Pick disease type C.";  
Hum. Genet. 105:10-16(1999).  
[10]  
RN VARIANTS NPC1 GLN-958 AND ALA-1007.  
RX MEDLINE=21313111; PubMed=11349231;  
RA Sun X., Marks D.L., Park W.D., Wheatley C.L., Puri V., O'Brien J.P., Kraft D.L., Lundquist P.A., Patterson M.C., Pagano R.B., Snow K.;  
RA "Niemann-Pick C variant detection by altered sphingolipid trafficking and correlation with mutations within a specific domain of NPC1.";  
Am. J. Hum. Genet. 68:1361-1372(2001).  
[11]  
RN VARIANTS NPC1 ALA-378; MET-950; ARG-992 AND THR-1061.  
RX MEDLINE=21313105; PubMed=11333381;  
RA Millat G., Marçais C., Tomasetto C., Chikh K., Pensom A.H., Harzer K., Wenger D.A., Ohno K., Vanier M.T.;  
RA "Niemann-Pick C1 disease: correlations between NPC1 mutations, levels of NPC1 protein, and phenotypes emphasize the functional significance of the putative sterol-sensing domain and of the cysteine-rich luminal loop.";  
Am. J. Hum. Genet. 68:1373-1385(2001).  
[12]  
RN VARIANTS NPC1 TYR-177; CYS-978 AND VAL-1035.  
RX MEDLINE=21372069; PubMed=11479732; DOI=10.1007/s004390100531;  
RA Ribeiro I., Marcao A., Amaral O., Sa Miranda M.C., Vanier M.T., Ribilat G.;  
RA "Niemann-Pick type C disease: NPC1 mutations associated with severe and mild cellular cholesterol trafficking alterations.";  
Hum. Genet. 109:24-32(2001).  
CC -!- FUNCTION: Involved in the intracellular trafficking of cholesterol. May play a role in vesicular trafficking in glia, a process that may be crucial for maintaining the structural and functional integrity of nerve terminals.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Found in late endosomes and lysosomes.  
CC -!- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain containing a di-leucine motif necessary for lysosomal targeting are critical for mobilization of cholesterol from lysosomes.  
CC -!- PTM: Glycosylated  
CC -!- DISEASE: Defects in NPC1 are the cause of Niemann-Pick disease type C1 (NPC1) [MIM:257220]. NPC1 is an autosomal recessive lipid storage disorder, which affects particularly the brain, liver and spleen, and which is characterized by lysosomal accumulation of low density lipoprotein derived cholesterol. Clinical features include variable hepatosplenomegaly and severe progressive neurological dysfunction such as ataxia, dystonia and dementia. The age of onset can vary from infancy to late adulthood.  
CC -!- DISEASE: Defects in NPC1 are the cause of Niemann-Pick disease type D (NPD) [MIM:257220]; also known as Niemann-Pick disease without sphingomyelinase deficiency, or Nova Scotian type. Because of evidence from biochemical changes, lack of complementation, and linkage mapping to the same chromosome site, NPD and NPC1 are considered to be allelic disorders.  
CC -!- SIMILARITY: Belongs to the patched family.  
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
-----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

DR EMBL; AF002020; AAB63982.1; -; mRNA.  
DR EMBL; AF157379; AAD48006.1; -; Genomic DNA.  
DR EMBL; AF157365; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157366; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157367; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157368; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157369; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157370; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157371; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157372; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157373; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157374; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157375; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157376; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157377; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF157378; AAD48006.1; JOINED; Genomic DNA.  
DR EMBL; AF338230; AAK25791.1; -; Genomic DNA.  
DR EMBL; AF123046; AAT28875.1; -; Genomic DNA.  
DR EMBL; AF123045; AAT28875.1; JOINED; Genomic DNA.  
DR EMBL; BC063302; AAH63302.1; -; mRNA.  
DR Ensembl; ENSG00000141458; Homo sapiens.  
DR HGNC; HGNC:7897; NPC1.  
DR MIM; 607623; -;  
DR MIM; 257220; -;  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR GO; GO:0005764; C:lysosome; TAS.  
DR GO; GO:0005624; C:membrane fraction; TAS.  
DR GO; GO:0005478; F:intracellular transporter activity; TAS.  
DR GO; GO:0015248; F:sterol transporter activity; TAS.  
DR GO; GO:0004888; F:transmembrane receptor activity; TAS.  
DR InterPro; IPR004765; NP\_C\_type.  
  
Query Match 34.8%; Score 2402.5; DB 1; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 5.8e-161;  
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;  
  
Qy 7 RGWLLWALLRLAQSEPTTIHQPGYCAFYDECG-----KNPELSGLMTLSNVSCLS 59  
Db 4 RGLALGALLLLCPAQVFSQ-----SCVYGEGLAYGDKRYNCEYSG-----46  
Qy 60 NTPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSLASISITFKALITRCPACSDN 119  
Db 47 --PPKPLPKDGYDLVQELCPGFFFG-NVSLCCDVQRQLQTLKDLQLPLQFLSRCPSCFYN 103  
Qy 120 FVNLIHCNTCSNQSLSLPIVTR-----VAQLGAGQLPAVAYEAFYQHSFABQSDVCSR 175  
Db 104 LLNLFCELTCSPRQSQFLNVTATEDYVDPTNQTNTKVKELQYVVGOSFANAMYNACRDV 163  
Qy 176 RVPAATAATLVGTWCVGVYGSALCNAQRWLNFGDGTGNGLAPLDTIT-----FHLLPQQA 228  
Db 164 EAPSSNDKALGLLGGKADA-CNATNWIETMKNKGQAPFTITPVFSDFPVH-----215  
Qy 229 VGGIQLPNEGVARCNESQGGDDVATCSQDCAASCAPAIARPQ-----ALDSTFYLG 279  
Db 216 ---GMEPMNATKGCDSVDEVTAPCSQDCSICVCGPKQPQPPPPAPWTILGLDAMYVIM 272  
Qy 280 QMPSGLVLIILLCISVFAVITLL-----VGRVAPARDKSKWDPKKGTSLS 326  
Db 273 WITVMAFLVFFGFAFFAVWCYRKRYFVSEYTPIDSNIAFSV-NASDRGE-----ASCC 324  
Qy 327 DKLSFSTHTLLGGFFQCGWGWASWPLTILVLSVIPVVALAAGLVFTETLTDPVELWSAP 386  
Db 325 DPTVSAAEFGCLRLRFTWGSFCVNRNPGCVIFFSFLVFTACSSGLVFRVVTNPNVDLNASP 384  
Qy 387 NSQARSKAFHDQHFQFFFTNQTQVILPAPNRSSRYRYSLLLPK-NFSGILDLLLELL 445  
Db 385 SSOARLEKEYFDQHFQFFFRTEQLIRAPLTDKHIYQYPPSGADVPFGPPLDIQLHQVL 444  
Qy 446 ELQERLRLQVWSPAEQARNLSLODICVAPLNPNSTLSYDCCINSLLQYFQNNRTLLTLTA 505  
Db 445 DLQIAIEN--ITASYDNETVTLLQDICLAPLSPYNT---NCTILSVLNFQNSHVSVDHKK 499  
Qy 506 NQTLMGQTSQVDWKDHFLYCANAPLTPKDGTLALALSCWADYGAPVFPFLAIGGYKGDYS 565

Db 500 GDDFF---VVADYHTEFLYCVRAPASLNDTSLHDPCLGTFGGVPFWLVLGGYDDQYN 556  
Qy EAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEMRAFORRMAGMFOVTFRTAERSLEDEI 625  
Db NATALVITFPVNNYNDTEKLQRAQAEKEFINFKYKN---PNLTISFAERSIEDEL 613  
Qy NRTTAEDLPATSYIVIFLYISALGSYSWSRWVDSKATLGLGGVAVVLGAVMAAMG 685  
Db NRESDSVFTWISYAIMFLYISALGHMKSRLLLVDSKVSGLGIAGILVLSVACSLG 673  
Qy FFSVLGRSSIVILQVFPFLVSLVAGADNIFIVLEYORLPRRPGEPREHIGRALGRVAP 745  
Db VFSYIGLPLTLIVIEPFLVAVGVNDIIFILVQAYQORDERLOGETDQOGLGRVIGRVP 733  
Qy SMLTCSLSEACFPGLGALTTPMPAVRTPALTSGLAVIDLDFLLQMSAFVALLSLDSKROEAS 805  
Db SMFLSSFSSEVAFPLGALSVMPPVHTFSLFAGLAVFIDFLQITCFVSLGLDQIKRQKN 793  
Qy RLDVCCCKPQOELPPPGQ-GEGLLGFQKAYAPFLHMTIRGVVLLFLALFVSLYSYM 864  
Db RLDIFCCVGAEDGTSSVOASECLFRFFKNSYSPLLKDWMPVIAIFVGVLSFSIAVL 853  
Qy CHLSVGLDQSLPKDSYLLDYFLFNRYFEVGAPYVFTVLGNFSSEAGMAICSSAG 924  
Db NKVDIGLDSLMPDDSYMDYFKSISQYLHAGPPVYFVLEEGHDYTSRQGNMVCGMG 913  
Qy CNNFSPTQKIOYATEPPEQSYLAIPASSWDDFDLWLT-SSCCRLVYISGPNKDKFCPT 983  
Db CNDLSVQOIFNAQLDNYTRIGFAPSWSWIDDFDWKPOSSCCRV---DNITDQPCNAS 970  
Qy VNSLNCNKCMST-MGSVRPSVEQPHKYLFWFLNDRPNIKCPKGLAAYSTSVN--LTS 1040  
Db VDPACVR-CRPLTPEGKQRPQGDPMRFLPMFLSDNPNFKCGKGHAAYSSAVNLLGH 1029  
Qy DGOVLASRFMAYHKPLKNSODYTEALPAARELANITADLRKVPGTDPAPPEVPPYITNV 1100  
Db GTRVGATYFMTYHTVLQTSADFTDALCKARLIASNVT-ETMGINGS--AYRVFPYSFVY 1086  
Qy FYEQYLITLPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSIVMILVDTVGFMALND 1160  
Db FYEQYLITIDTTFNLGVSLGALFVTVMLLGCLESAVJMCATIAMVLMFVGMWLMG 1146  
Qy ISYNAVSLINLVSAVMSVEFVSHITRSEAFISTKPTWLERAKEATISMGSAVPAVAMTN 1220  
Db ISLNAVSLNVLVMSCGISVEFCSHITRAFTVSMKGSVEREAEALAHMGSSEVSGITLTK 1206  
Qy LPGILVLGLAKAQLIOIFPRLMLTLTLGLHLGLFPLVILSYVGDVNPA 1272  
Db FGGIVVLAFAKSQIFQIFFRMYLAMVLLGATHGLIFLPLVLLSYIGPSVYKA 1258

## RESULT 10

Q5GRI HUMAN  
ID Q5GRI\_HUMAN PRELIMINARY; PRT; 1289 AA.  
AC Q5GRI;  
DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Niemann-Pick disease, type CI variant (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RC NUCLEOTIDE SEQUENCE.  
RP TISSUE=Brain;  
RA Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,  
RA Ohara O., Nagase T., Kikuno F.R.;  
RT "None title."  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB209048; BAD92285.1; -; -.

FT NON TER 1  
SQ SEQUENCE 1289 AA; 143230 MW; 663DE27C13FD801F CRC64;  
Query Match 34.8%; Score 2402; DB 2; Length 1289;  
Best Local Similarity 39.3%; Pred. No. 6.3e-161;  
Matches 525; Conservative 239; Mismatches 452; Indels 106; Gaps 26;  
Qy 2 AEAGLRGWLWA-----LLRLAOSPEVTTIHQPGYCAPYDECG-----KNPELSG 47  
Db 3 ASAAIRGHSTARGALGLLGLLLCPAQVFSQ-----SCVYGECCGAYGKRYNCYSYG 57  
Qy 48 SMLTNSVCLSNTPARKITGDHLLILQKICPLRYTGPNTOACCSAKQLVSLASLITK 107  
Db 58 -----PPKPLPKDGYDLVQELCPGFFG-NVSLCCDVRQLQTLKDNQLPL 102  
Qy 108 ALITRCPACSDNFVNLCHNTCSPNOSLFINVT-VAQLGAGQLPAVVAEAPQHS 163  
Db 103 QFUSRCPCFPYNLNLFCELTSPSQSOFNLNVTATEDYDVPVTNQTNTVKEQLQYVVGQS 162  
Qy 164 FARQSYDSCSRVRVPAATAITLAVGTCGVTSALCNAQRWLNFGDGTGNGLAPLDT--- 219  
Db 163 PANAMYNACEDVEAPSNDKALGLCGKADA-CNATNWIEMPNKDNQGOAPPTITPVFS 221  
Qy 220 -PHLLBPGQAVSGIOPLNEGVARCNESQGDVATCSQDCAASCAPAIAPQ----- 270  
Db 222 DFPV-----RGMEPMNATKGCDESVDVETAPCSCDCSIVCGPKPQPPPPAPWTI 273  
Qy 271 -ALDSTFYLGQPGSLVLIILLCSEFVAVTILL-----VGRVAPARDKSKM 316  
Db 274 LGLDAMYIMWTYMAFLVFFGAFVAVMCYRKRYFVSEYTPIDSNIAFSV-NASDKGE- 331  
Qy 317 VDPKGTSLSDKLSFSTHTLLGQFGQGTWASWPLTILVLVSVIPVALAAGLVFTLT 376  
Db 332 -----ASCDDPVSAAPEGCLRLPFRWGSFCVRNPGCVIFFSLVITACSSGLVFRVT 385  
Qy 377 TDPVELWSPNSQARSEKAFHDQHPFPFRTNQVILTAPNRSYRYSDDLGLPK-NFSGI 435  
Db 386 TNPVDLWSAPSSQARLEKEYFDQHGFFFRTEQLIIRAPLTDKHIQYPSGADVPEGPP 445  
Qy 436 LLDLLELLELELQERLHLQVMSPEAQRNLSODICVAPLNPDNTSLYDCCINSLOYFQ 495  
Db 446 LDIQLHLVDLQIAIEN--ITASDYDNETVTTLQDICLAPLSYNT--NCTILSVLNYFQ 500  
Qy 496 NNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGITALALSCMADYAGVAPPELA 555  
Db 501 NHSVLDHKKGDDFF---VYADYHTEFLYCVRAPASLNDTSLHDPCLGTFGGVPFWLV 557  
Qy 556 IGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEMRAFORRMAGMFOVTF 615  
Db 558 LGGYDDQNNATALVITFPVNNYNDTEKLQRAQAEKEFINFKYKN---PNLTISF 614  
Qy 616 TAERSLEDEINRTAEDLPATSYIVIFLYISALGSYSWSRWVDSKATLGLGGVAV 675  
Db 615 TAERSIEDLNRSDSDVFTVVISAIMFLYISALGHMKSRLLLVDSKVSGLGIAGILI 674  
Qy 676 VLGAVMAAMGFFSYLGIRSLVILQVFPFLVSLVAGADNIFIVLEYORLPRRPGEPREH 735  
Db 675 VLSVACSLGVSYGLPLTLIVIEVIFPLVAVGVNDIIFILVQAYQORDERLOGETDQO 734  
Qy 736 IGRALGRVAPSMLLCSLSEACFPGLGALTTPMPAVRTPALTSGLAVIDLDFLLQMSAFV 795  
Db 735 LGRVLGEVAPSMFLSSFSSEVAFPLGALSVMPPVHTFSLFAGLAVFIDFLQITCFVSL 794  
Qy 796 SLDSKROEASRLDVCCCKPQOELPPPGQ-GEGLLGFQKAYAPFLHMTIRGVVLLFL 854  
Db 795 GLDIKROEKRLDIFCCVGAEDGTSSVOASECLFRFFKNSYSPLLKDWMPVIAIFV 854  
Qy 855 ALFGVSLYSMCHISVGLDQELAPKDSYLLDYFLFNRYFEVGAPYVFTVLGNFSSEA 914  
Db 855 GVLSTFSIAVLNKVDIGLDSLMPDDSYMDYFKSISQYLHAGPPVYFVLEEGHDYTSK 914  
Qy 915 GMAICSSAGCNFSPTQKIOYATEPPEQSYLAIPASSWDDFDLWLT-SSCCRLVYISG 973

Db 915 GQNMVCGMGCCNNDLSVQOIFNAAQLDNVTRIGFAPSSWIDDYFDWVKPQSSCCRV---D 971  
QY 974 PKNDKFCSTVNSLNCNKCMSIT-MGSVRSEVQFHKYLPWFLNDRNINIKPKGGLAAY 1032  
Db 972 NITDQFCNASVDPACVR-CRLPTPEGRQPGQDFMRFLPMFLSDNPNPKCGKGHAAY 1030  
QY 1033 STSVN--LTSQGVASRFMAVHKPKNSQDYTEALRAARELAANITADLRKVRGTPDAF 1090  
Db 1031 SSVANILGHTRVGATFTMYTHVLTQTSADFIDALKARLIASNVT-ETWINGS--AY 1087  
QY 1091 EVFPYTIITNVFEQYLTITPEGLFMLSCLVPTFAVSCLLGLDLSRGLNLLSTVMILV 1150  
Db 1088 RVFPYSVFVFEYQYLTITDDTFNLGVSGLGAIPLVTWVLGCELWSAVIMCATIAMVLV 1147  
QY 1151 DTVGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPFAISTKPTWLERAKEATISMG 1210  
Db 1148 NMFQVWMLGWSLNAVSLVNLVMSGSGISVEFCSHITRAFTVSMKGSRYERAEALAHMS 1207  
QY 1211 AVFAGVAMTNLPGILVLGLAKAQLIQIPFFRLNLLITLGLLHGLVPLVPLVILSVYGP 1270  
Db 1208 SVFSGITUTKFGGIVVLAFAKSQIFQIFRYRMLANVLGATHGLIFLPLVLSYIGPSVN 1267  
QY 1271 PA 1272  
Db 1268 KA 1269  
RESULT 11  
Q7TMD4 MOUSE  
ID Q7TMD4 MOUSE PRELIMINARY; PRT; 1277 AA.  
AC Q7TMD4.  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 10-MAY-2005 (TREMELrel. 30, Last annotation update)  
DE Nienann Pick type C1.  
GN Name=Npcl;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Hellon E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Brain;

RA Strausberg R.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC052437.1; AAH52437.1; -; mRNA.  
DR EMBL; BC054639.1; AAH54639.1; -; mRNA.  
DR GO; GO:0006897; P: endocytosis; IMP.  
DR InterPro; IPR004765; NP\_C type.  
DR InterPro; IPR003392; Patched.  
DR InterPro; IPR000731; SSD\_5TM.  
DR Pfam; PFO2460; Patched; 1.  
DR TIGRFAMs; TIGR00917; 2A060601; 1.  
DR PROSITE; PS0156; SSD; 1.  
SQ SEQUENCE 1277 AA; 142882 MW; 3B42230AAC8E564E CRC64;  
Query Match 34.7%; Score 2400; DB 2; Length 1277;  
Best Local Similarity 40.4%; Pred. No. 8.7e-161;  
Matches 528; Conservative 233; Mismatches 443; Indels 104; Gaps 27;  
QY 11 LWALLLLAASEPPTTHIQPGYCAFYDECGNPELSGLMTLSNVSCLSNTPARKITGDH 70  
Db 8 LGLLLLLLCPAQVESQ-----SCVWYGEQG---IATGD---KEYNCKYSGPPKPLPKDG 55  
QY 71 LILQKTCPLRYTGNTPQACSAKQVLSLEASISITKALLTRCPACSDNFVNLHCHNTCS 130  
Db 56 YDLVQELCPGLFF-DNVSLLCCDIQQLQTLXSNLQLPLQFLSRCPSCFYNLMTLFCELTCS 114  
QY 131 PNQSLFNTVTRVAQLGAGQLPA----VVAVEAFVQHSFAESYDSCSRVRVPAATLAVG 186  
Db 115 PHOSQFLNVTATEDYDPKTOENKTNVKELEYFVGQSFANAMYNACRDVEAPSNKALG 174  
QY 187 TMCVGYSGALCNAQRWLNFOQDTGNGLAPLDI-----TFHLEPGQAVGSGIQPLNEGVA 241  
Db 175 LLCGRDARA-CNATNWIEYMFENKNGQAPFTIIPVFSDSLTL-----GMEPWRNATK 225  
QY 242 RCNESQGDVATCSCQCAASC-----PATARQALDSTF-----VILQMPG 283  
Db 226 GCNESVDEVGTGPCSCQDCSICVCGPKQPPPPPPMPRIWGLDAMVIMWVTVVAELFVFFG 285  
QY 284 SLVLIILLCSPAVVVTILLGFRVAPARDKSKMDPKGTSLSDKLSFSTHTLLGQFPQG 343  
Db 286 ALLAVWCHRRIRYFVSEYTPIDSNJAFSVNS-----DKGEASCCDPLGAFDDCRRMPTK 341  
QY 344 WGTWVASWPLTILVSVIPVVALAAGLVFTLTDTPVELMSAPNSQARSEKAFDQHPG 403  
Db 342 WGAFCVRNPTCIIFFSLAFITVCSSGLVFVQVTTNPVELMSAPHQARLEKEYDKHFGP 401  
QY 404 FFRNQVILFAPNRSRYDSLLGPK-NFSGIILDLLELLELQERLRLHQWSPSEAQ 462  
Db 402 FFRTEQLIQAPNTSVHIYEPYPAGADVPPFGLPKELHQLVLDQ-----IAIESITAS 456  
QY 463 RN---ISLQDICVAPLNPDTLSYDCCINSLLQVFNQNRLLLTLLLTANQTLMGQTSQV--- 516  
Db 457 YNNTVTIQLDLCVAPLSFYNK---NCTIMSVNLTQNSHAVL-----DSQVGD 502  
QY 517 -----DWKDFLYCANAPLTFKDGTLALSCMADYGAPVPFPFLAIGGYKGYGEAEAL 571  
Db 503 FYIYADVHTHFLYCVRAPASLNDTSLHLGPGCLGTGGVPFVPLVLGGYDQYNNNATLV 562  
QY 572 MTPSLNYPAGDPRLAQAQWEEAFLEEMRAFORMMAGMFOVTTAERSLEDEINRTAE 631  
Db 563 ITFPVNNYNDTERLQRAWEKEFISFVKYKN---PNLTISPTAERSIEDELNRNSNS 619  
QY 632 DLPIFATSYIVIFLYISALGYSYSGRVMVDKATGLGAVVVLGAVMAAGFFSYLG 691  
Db 620 DVFTVILSVVMFLYISALGHIOCSRLLVDSKISLGIAGIILVLSVACSLGIFSYMG 679  
QY 692 IRSSLVILQVVPFLVSVGADNIFIVLEYQRLPRRPFEPREHVHIGRALGRVASMILCS 751  
Db 680 MPLTLIVIEVIFLVLAUGVDNIFILVQTYORDERLQEBETLDQQLGRILGEVAPTMFLSS 739  
QY 752 LSEAIACFLGNLTMPAVRTFALTSLGLAVILDFLLQMSAFVALLSLQSKRQESRLDVCC 811  
Db 740 FSETSAFFFGALSSMPAVHTFSLFAGMAVLIDFLQITCFVSLGLGLDKRKKNHLDILC 799

Qy	812	CVKQELPPPOG-----EGLLGFFQKAYAPFLHWHITRGVLLFLALFGVSYISMCHI	867
Db	800	CVRGAD--DQGGSHASESYLFRFFKNYFAPLLKDWLRPIWAVFVGVLSFSAVVNVKV	856
Qy	868	SVGLDQELALPKDSYLLDYFLFLNRYFEVGAFFVFTTLGVNFSSEAGMNAICSSAGCNV	927
Db	857	DIGLDQSLSPNDSYVIDYFKSLAQYLHSGPPVTVLEEGYNSRSRGQNNMVCQGMGCDN	916
Qy	928	FSFTQKIQYATEPPEQSYLAIPASSWVDDFIDMLTP--SSCCRLYISGPNKDKFCPSTVNS	986
Db	917	DSLVOQIFNAELDTYTRVGFAPSSWIDDYFDWVPOSSCCRLY---NVTHQFCNASVMD	973
Qy	987	LNCLKNCSIT-MGSVRPSVEQFHYLPWFILNDAPNICKPKGGLAAYSTVNLTSQGVLT	1045
Db	974	PTCVR--CRPLTEPGKQPOGKEFKFLFMFLSDNPNPKCGKGHAAYGSAVNIVGDDTYI	1032
Qy	1046	-ASRFMYHKLKNSQDYTEALRAARELANITADLRKVPCTOPAFEPVPTITNVFEQ	1104
Db	1033	GATYFMYTHILKTSADYTDAMKARLIASNITETMRS--KGSD--YRVFPYSVPVFEQ	1089
Qy	1105	YLTLPEGLFMLSCLVPTFAVSCLLGLDLRLSGLLNLSTVMILVDTGVGFMALWDISYN	1164
Db	1090	YLIIDDTIFNLVSLGSIFLTVLVGLCELSAVINCIITAMLVNMFGVWMLGSLN	1149
Qy	1165	AVSLINLVAGMSVFVSHITRSFAISTPTWLERAKEATINSGSAVFAVAMTNLPGI	1224
Db	1150	AVSLNVLNMGSGISVEPCSHITRAFTMSTKGRSVRAEALAHMGSSVFGSITLTKFGGI	1209
Qy	1225	LVLGLAKAQLIQIFFRELNLITLGLLHGLVFLPVLTLSVVGDDVNP	1272
Db	1210	VVLAFAKSQIFEIFFRMYLAMVLLGATHGLIFLVLVLLSYIGPSVNKA	1257
RESULT 12			
ID	Q9N0Q0_FELCA	PRELIMINARY;	PRT; 1276 AA.
AC	Q9N0Q0;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Niemann-Pick type C1 disease protein.		
GN	Name=NPCI;		
OS	Felis silvestris catus (Cat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;		
OC	Felis.		
OX	NCBI_TaxID=9685;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	TISSUE=Kidney;		
RA	Murakawa M., Freeman M.W.;		
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AP258783; AF272187.1; -, mRNA.		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0005764; C:lysosome; IEA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0008158; F:hedgehog receptor activity; IEA.		
DR	GO; GO:0030301; P:cholesterol transport; IEA.		
DR	InterPro; IPR004765; NP_C_type.		
DR	InterPro; IPR003392; Patched.		
DR	InterPro; IPR000731; SSD_5TM.		
DR	Pfam; PF02460; Patched; 1.		
DR	TIGRFAMs; TIGR00917; 2A060601; 1.		
DR	PROSITE; PS50156; SSD; 1.		
SQ	SEQUENCE 1276 AA; 141739 MW; CB9B7A0261691256 CRC64;		
Query Match			
Best Local Similarity 40.6%; Pred. No. 1e-160;			
Matches 534; Conservative 232; Mismatches 462; Indels 86; Gaps 29;			
Qy	7	RGWLLWALLRLAQSPYTIHQPGYCAFVDEGKKNPELSGMTVNSVCSLNTNPARKI	66
Db	4	RGPALGLLLLLC---PAQVLAQS--CIWYGECS---IASGD-----KRYNCKYSGPPKPL	51

Qy	67	TGDLHLLQKICPRLYTGPNTQACCSAKQVLVSLEASISITKALITRCPACSDNFVNLHCH	126
Db	52	PKGYDLVQELCPGFFP-DNVSCLCCDVQQLQTLKDNLQLPLQFLSRCPCFYNLVNLFCE	110
Qy	127	NTCSPNQSLPINVTR-----VAQLGAGQLPAVAYEAPYQHSFPAEQSYDCSRVRVPAAT	182
Db	111	LTCSPRQSQFLNVTATEDYVDPVNTQTKTNVKELQYYIGESFANAMYNACRDVEAPSND	170
Qy	183	LAVTMCVGVYSALCNAQRWLNFGQDTNGCLAPLDIT--PHLLEPGQAVSGIQPLNEGV	240
Db	171	KALGLICGKDAEA-CNATNWIYFNSKDNQQAFTTIPFSDLPT-----HGMPEMNAT	224
Qy	241	ARCNESGDDVATCSCQDCAASC-----PAIARPOALDSTFYLGQMPGSLVLIIIL	291
Db	225	KGCDSEVDEVTGFCSCQDCSIVCGPKPQPPPPVPMRILGLDAMYVIMWTYMAFLVFF	284
Qy	292	CSVFA-----VTVILLVGRVA---PARDKSKMVDPKGTGSLDKLSFSFTHLLGQ	339
Db	285	GAPFALWCYRKRYFVSEYTPIDSNIAFSVNANDRGE-----ASCCDALGAAPGCLRR	337
Qy	340	FFQGMGTWVASWPLTILVLSVIPVALAAGLVFETLTDPELWLSAPNSOARSEKAFHQ	399
Db	338	LFSQWGSFCVRNPGPIIFFSLAFIAACSSGLVFRVVTNPVDLWSAPSSQARLEKEYFT	397
Qy	400	HFGPFRFTNQVILTAPNRSRYSDSLLLGPK-NFSGILDLDLLELELELERLRHLQVMS	458
Db	398	HFGPFRFTQELIIQAHPHTSAHTYQYPGSDVPFGPPLDLAILHOVLDTQTEN--ITA	455
Qy	459	PEAQRNISLDICYAPLNPDNTSLYDCCINSLLQYFQNNRTLLLTANOTLMGQTSQV--	516
Db	456	SYNNETVTLLQDLCVAPLSPYNK---NCTILSVLNYFQNSHML-----DHEIGDDPFVA	507
Qy	517	DKKDHELYCANAPLTFKDGTTALALSCHADYGAVFPFLAIGGYKGDYSABEALIMTFSL	576
Db	508	DYHTHLLYCVRAPASLNDYSLLDHPCLGTFGGVPFWLVGGYDDQNNYNATALVITFPV	567
Qy	577	NNYPAGDPRLAQAKLWEEAFLEEMRAFQRRMAGMFVFTTAERSLEDEINRTTAEDLPIF	636
Db	568	NNYNDTERLQAHVWEKEFINVKNYKN---PNLTISFTTERSIEDELNRSGNDIFTV	624
Qy	637	ATSYIVIFYLSIALGYSYSSWRMVDSKATLGLGVAVVLGAVMAAMGFYSYLGIRSSL	696
Db	625	IISYATMFLYISIALGHIKSCSRLVDSKISLAGILIVLSKACSLGIFSVMGVLPLT	684
Qy	697	VILQVPELVLSVGADNIEIFVLEYQRLPRRGPCEPREVHIGRALGVAPSMILCSSEAI	756
Db	685	IVIEVIPFLVAVGVNDIIFILVQYQYQDERLHGETLDQQLGRVIGSVAPSMFLSSFSEAV	744
Qy	757	CFPLGALTWPAVRTFALTSGLAVIDLLOMSAFVALLSLDSKROEASRLDYCCCKVPQ	816
Db	745	APFLGALSKMPAVHTFSLFAGWAVLIDFLQITCFVSLGLGLDIKROEKRLDLVCCVRS	804
Qy	817	ELPPPPQ-GEGLLLGFFQKAYAPFLHWHITRGVLLFLALFGVSYISMCHISVGLDQEL	875
Db	805	EDGTSVQASESCLFRLFKHSYSPLLKDWMRPIVIAIFVGVLSFSAVLNKKVEIGLDQSL	864
Qy	876	ALPKDSYLLDYFLFLNRYFEVGAFFVFTTLGVNFSSEAGMNAICSSAGCNPSFTQKIQ	935
Db	865	SMPDDSYNMDYFKSL-KYLHAGPPVYFVLEEGHDYTSLKQNNMVCQGMGCDNLSVQQIF	923
Qy	936	YATEPPEQSYLAIPASSWVDDFIDMLTP--SSCCRLYISGPNKDKFCPSTVNSLNCNKCM	994
Db	924	NAQLDSYTRIGFAPSSWIDDYFDWVKPOSSCCRVNS---TDRFCNASVVDPAQICR-CR	979
Qy	995	SITM-GSVRPSVEQFHYLPWFILNDAPNICKPKGGLAAYSTSVN-LTSDGQVLASPMAY	1052
Db	980	PLTQEGKQRPQGGDFMRFLPMFLSDNPNPKCGKGHAAYSSAVNLGNDTGVGATVFTY	1039
Qy	1053	HKPLKNSQDYTEALRAARELANITADLRKVPCTDPA-FEVEPYITNVFYEQYLILPE	1111
Db	1040	HTVLQTSADFTDAMRKANLIASNIT-----KTMGLEGSNRYRVFPYSVYFYEQYLTIID	1095







10-MAY-2005 (Rel. 47, Last annotation update)  
 DT Niemann-Pick C1 protein precursor.  
 GN Name=Npci;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=97362324; PubMed=9211850; DOI=10.1126/science.277.5323.232;  
 RA Loftus S.K., Morris J.A., Carstea E.D., Gu J.Z., Cummings C.,  
 RA Brown A., Ellison J., Ohno K., Rosenfeld M.A., Tagle D.A.,  
 RA Pentchev P.G., Pavan W.J.;  
 RT "Murine model of Niemann-Pick C disease: mutation in a cholesterol  
 homeostasis gene.";  
 RL Science 277:232-235 (1997).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=99145606; PubMed=9990080; DOI=10.1073/pnas.96.4.1657;  
 RA Patel S.C., Suresh S., Kumar U., Hu C.Y., Cooney A.,  
 RA Blanchette-Mackie E.J., Neufeld E.B., Patel R.C., Brady R.O.,  
 RA Patel Y.C., Pentchev P.G., Ong W.-Y.;  
 RT "Localization of Niemann-Pick C1 protein in astrocytes: implications  
 for neuronal degeneration in Niemann-Pick type C disease.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:1657-1662 (1999).  
 CC -!- FUNCTION: Involved in the intracellular trafficking of  
 cholesterol. May play a role in vesicular trafficking in glia, a  
 process that may be crucial for maintaining the structural and  
 functional integrity of nerve terminals.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Found in late  
 endosomes and lysosomes.  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in perisynaptic  
 astrocytic glial processes. Also expressed in heart, spleen, lung,  
 liver, skeletal muscle, kidney, testis.  
 CC -!- INDUCTION: Activated by the drugs progesterone and U-18666A which  
 block cholesterol transport out of lysosomes and by the  
 lysosomotropic agent NH4Cl.  
 CC -!- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain  
 containing a di-leucine motif necessary for lysosomal targeting  
 are critical for mobilization of cholesterol from lysosomes.  
 CC -!- SIMILARITY: Belongs to the patched family.  
 CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 CC -----  
 DR EMBL; AF003348; AAB63372.1; -; mRNA.  
 DR EMBL; AF003349; AAB63373.1; -; Genomic DNA.  
 DR FIR; T30188; T30188.  
 DR Ensembl; ENSMUSG0000024413; Mus musculus.  
 DR GO; GO:006897; P: endocytosis; IMP.  
 DR GO; GO:004765; NP C type.  
 DR InterPro; IPR004392; Patched.  
 DR InterPro; IPR000731; SSD\_5TM.  
 DR Pfam; PF02460; Patched; 1.  
 DR TIGRFAM; TIGR00917; 2A060601; 1.  
 DR PROSITE; PS0156; SSD; 1.  
 KW Glycoprotein; Lysosome; Signal; Transmembrane.  
 FT SIGNAL 1 23 Potential.  
 FT CHAIN 24 1278 Niemann-Pick C1 protein.  
 FT TRANSMEM 271 291 Potential.  
 FT TRANSMEM 352 372 Potential.  
 FT TRANSMEM 623 643 Potential.  
 FT TRANSMEM 655 675 Potential.  
 FT TRANSMEM 685 705 Potential.  
 FT TRANSMEM 761 781 Potential.  
 FT TRANSMEM 834 854 Potential.  
 FT TRANSMEM 1099 1119 Potential.

TRANSMEM 1125 1145 Potential.  
 FT TRANSMEM 1196 1216 Potential.  
 FT TRANSMEM 1228 1248 Potential.  
 FT DOMAIN 621 786 SSD.  
 FT MOTIF 1275 1278 Di-leucine motif.  
 FT COMPTIAS 250 260 Poly-Pro.  
 FT CARBOHYD 71 71 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 123 123 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 138 138 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 186 186 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 223 223 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 229 229 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 415 415 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 460 460 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 479 479 N-linked (GlcNAc..)(Potential).  
 FT CARBOHYD 525 525 N-linked (GlcNAc..)(Potential).  
 FT CONFLICT 474 474 S -> F (in Ref. 1; AAB63373).  
 FT CONFLICT 479 479 N -> D (in Ref. 1; AAB63373).  
 SQ SEQUENCE 1278 AA; 142889 MW; 43C71CE47D283674 CRC64;  
 Query Match 34.5%; Score 2385; DB 1; Length 1278;  
 Best Local Similarity 40.2%; Pred. No. 1e-159;  
 Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;  
 QY 14 LLRLAQSBPYTHIQPGYCAFYDECGKNPELGSGLMTLSNVCLSNTPARKITGDHLIL 73  
 DB 12 LLLLLCPAQVFSQ-----SCVWYGECC---IATGD---KRYNCKYSGPKPLPKDGYDL 59  
 QY 74 LQKICRLTYGPNTOACCSAKQLVSLASISITKALLTRCPACDPNFVNLIHCHNTCSPNQ 133  
 DB 60 VQELCPGLFF-DNVSLCCDIQQLTKLSNLQLPLQLSCPCFCYFNLMTLFCELTCSPHQ 118  
 QY 134 SLFINTVRAQLGAGQLPA-----VVAEAFYQHSFAEQSYDCSRVVRPAANTLAVTMC 189  
 DB 119 SQFLNVTATEDYDPKTPENKTNVKELYVVGOSFANAMYNACRDVEAPSSNEKALGLIC 178  
 QY 190 GYVGSALCNAQRWLNFGDTGNGLAPLDI-----TFHLLPEPQAVSGSIGIPLNEGVARCN 244  
 DB 179 GRDARA-CHNTNWIWFKNQNGQAPFTIIPVPSDSLIL-----GMEPMRNTKGCN 229  
 QY 245 ESGGDVATCSQDCAASC-----PAIARPAQALDSTF-----YLGQMPGSLV 286  
 DB 230 ESDVEVTGPCQDCSIVCGPKPQPPPPMPWIRWGLDAMYIMWVTVYVAFVFFGALL 289  
 QY 287 LIILLCVNAVITLLVGRFVAPARDKSKMVDPKGTSISDKLSFSTHTLLGFFQGWGT 346  
 DB 290 AVWCHRRRYFVSEYTPIDNSIAFVNSV-----DKGEASCCDPLGAADFDCLERMFTKWA 345  
 QY 347 WVASWPLTILVLSVIPWALAGLVFTLTDPVELWSAPNSQARSEKAFHDQHFQFFR 406  
 DB 346 FCVRNPTCIFFSLAFITVCSGLVFVQVTTNPVELWSAPHSQARLEKEYFDKHFQFFR 405  
 QY 407 TNQVILTAPNRSRYSDLSLLGPK-NFSGILDLDLLELLELLELLELLELLELLELLELLEL 464  
 DB 406 TEQLITQAPNTSVHIEPYEPAGADVFPGLNKEILHQLVNLQ-----IAIESITASYNN 460  
 QY 465 --ISLQDICVAPLNPDNTSLYDCCINSLLQYQNNRTLLLTANQTLMGTSQV----- 516  
 DB 461 ETVTLDQICVAPLSPYNK---NCTINSLVNLVFNQSHAVL-----DSQVGDDFYI 506  
 QY 517 --DWKDFLYCANAPLTFKDGFTALALSCMADYGAPVFPFPLAIGGYKGYKDYSEALIMTF 574  
 DB 507 YADYTHFLYCYRAPASLNDTSLHGPCGLGTGGPVFPWLVLGGYDDQYNNNTALVITF 566  
 QY 575 SLNNYPAGDPRLAQAKLWEEAFLEENRAFQRRMAGMFQVTTFAERSLEDEINRTTAEDLP 634  
 DB 567 PVNNYNDTERLQRAWAWEKEFISFVKNYK---PNLTISPTAERSIEDENRESNDVF 623  
 QY 635 IFATSYIVFLYISALGYSYSSRWVWDKATLGLGVAVLGVAVLGVAVLGVAVLGVAVLGVAVL 694  
 DB 624 TVIISVVMFLYISLALGHQICSRLLVDSKISLGIAGILIVLSSVACSLGIFSYNGMPL 693  
 QY 695 SLVILQWVFLVLSVGADNIFIFVLEYQRLPRRPGEPREHVHIGRALGRVAPSMILCLSLSE 754



Db	994	DFMRFLPFLSDNPNPKCGKHRAAASAAVNI	LDNGTRVGATYFMTYHTVLQTSADFIDA	1053
Qy	1066	LRAARELAANITADLRKVP	GTDPAFEFVPTITNVFYEQYLTILPEGLFMLSCLVPTPA	1125
Db	1054	MEKARLIASNITRTMNOQGD--	HRVFPYSVFVFEQYLTWIDDTIFNLSVSLGAIFL	1110
Qy	1126	VSCLLGLDLRSGLLNLLSI	VMILVDTVGFMAWDISYNAVSLINLVSAYGMSVEFVSHI	1185
Db	1111	VAVVLLGCELWSAVIMCATI	AMILVNMFGVNLWGISLNAVSLVNLVMSCGISVEFCSHI	1170
Qy	1186	TRSFALSTKPTWLERAKEAT	ISMGSAVEFAGVAMTNLPGLVILGLAKAQLIQI	1245
Db	1171	TRAPTSTKGRVERAEEALSH	MGSSVFSGITLTGFGIIVLAPAKSQIFQIFFRMYLA	1230
Qy	1246	ITLLGLHLGLVFLPVLIS	VYVGPDPVNPALAL	1275
Db	1231	MVLGATHGLIFLPVLLSYI	GPSINKAKSL	1260

Search completed: April 7, 2006, 19:21:05  
Job time : 274 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:21:19 ; Search time 52 Seconds  
(without alignments)  
2117.769 Million cell updates/sec

Title: US-10-736-769-4  
Perfect score: 6909  
Sequence: 1 MAEAGLRGWLWALLRLAQ.....GSIKAGAGISNPLPNNGRQF 1332

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2403	34.8	1318	2	US-09-949-016-10152
2	2402.5	34.8	1278	2	US-09-462-136-2
3	2385	34.5	1319	2	US-09-462-136-4
4	1329	19.2	1170	2	US-09-462-136-6
5	1078	15.6	1296	2	US-09-462-136-9
6	651.5	9.4	1447	1	US-08-540-406-19
7	651.5	9.4	1447	2	US-08-656-055-19
8	651.5	9.4	1447	2	US-08-954-668-19
9	651.5	9.4	1447	2	US-09-268-140-5
10	651.5	9.4	1447	2	US-08-918-658-19
11	651.5	9.4	1447	2	US-09-724-631-19
12	651.5	9.4	1447	2	US-08-954-701A-19
13	651.5	9.4	1447	2	US-09-807-007-6
14	651.5	9.4	1447	2	US-09-754-032-19
15	651.5	9.4	1447	2	US-08-916-140-19
16	651.5	9.4	1447	4	PCT-US95-13233-19
17	646.5	9.4	1434	1	US-08-540-406-10
18	646.5	9.4	1434	2	US-08-656-055-10
19	646.5	9.4	1434	2	US-08-954-668-10
20	646.5	9.4	1434	2	US-08-918-658-10
21	646.5	9.4	1434	2	US-09-724-631-10
22	646.5	9.4	1434	2	US-08-954-701A-10
23	646.5	9.4	1434	2	US-09-754-032-10
24	646.5	9.4	1434	2	US-08-916-140-10
25	646.5	9.4	1434	4	PCT-US95-13233-10
26	639.5	9.3	642	2	US-09-248-796A-16161
27	628	9.1	1296	2	US-08-857-636-60

28	584	8.5	1203	2	US-09-293-505-2	Sequence 2, Appli
29	584	8.5	1203	2	US-09-060-939A-2	Sequence 2, Appli
30	583.5	8.4	1182	2	US-09-293-505-7	Sequence 7, Appli
31	583.5	8.4	1182	2	US-09-060-939A-7	Sequence 7, Appli
32	581.5	8.4	531	2	US-09-248-796A-16160	Sequence 16160, A
33	580	8.4	1203	2	US-09-207-857-2	Sequence 2, Appli
34	580	8.4	1203	2	US-09-909-280A-2	Sequence 2, Appli
35	579	8.4	1146	2	US-09-807-007-1	Sequence 1, Appli
36	526.5	7.6	1311	1	US-08-540-406-4	Sequence 4, Appli
37	526.5	7.6	1311	2	US-08-656-055-4	Sequence 4, Appli
38	526.5	7.6	1311	2	US-08-954-668-4	Sequence 4, Appli
39	526.5	7.6	1311	2	US-08-918-658-4	Sequence 4, Appli
40	526.5	7.6	1311	2	US-09-724-631-4	Sequence 4, Appli
41	526.5	7.6	1311	2	US-08-954-701A-4	Sequence 4, Appli
42	526.5	7.6	1311	2	US-09-754-032-4	Sequence 4, Appli
43	526.5	7.6	1311	2	US-08-916-140-4	Sequence 4, Appli
44	526.5	7.6	1311	4	PCT-US95-13233-4	Sequence 4, Appli
45	474	6.9	1286	2	US-09-268-140-3	Sequence 3, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-949-016-10152  
; Sequence 10152, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10152  
; LENGTH: 1318  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10152

Query Match		34.8%;	Score 2403;	DB 2;	Length 1318;
Best Local Similarity		39.7%;	Pred. No. 3.9e-218;		
Matches 525;		Conservative 238;	Mismatches 451;	Indels 110;	Gaps 26;
QY	2	AEAGLRGWLWA-----LLRLAQSEPYTHIQPGYCAFYDECG-----KNPELSG	47		
Db	32	ASAAALRGHSMTARGALGLLLLLCPAQVFSQ-----SCWYGEAGIAYGDKRYNCEYSG	86		
QY	48	SLMTLSNVSCLSNTPARKITGDHLLILQIKIPRLTYGNTQACCSAKOLVSEASLSITK	107		
Db	87	-----PPKPLPKDGYDLVOELCGGFFG-NVSLCCDVRQLQTKNKLQDPL	131		
QY	108	ALLTRCPACSNFVNLHCHNNTCSFNQSLFINVTR-----VAQLGAGQLPAVVAYEAFYQHS	163		
Db	132	QFLSRCPSCFYNNLLNLPCELTCSPRSQSLNVTATEDYDVPVTNQTKNVKELQYVYVQS	191		
QY	164	PAEQSYDSCSRVVRPAATLAVGTWCGYVGSALCNAORWLNFOQDGTGNGLAPLDTIT-----	219		
Db	192	FANAMYNACRDVEAPSSNDKALGLLCGKDADA-CNATNWIIEYMFNFKONGQAPFTTTPVFS	250		
QY	220	---PHLEPGQAVSGIOPLENEGVARNESGDDVATCSCODCAASCPAIRPQ-----	270		
Db	251	DFPVH-----GMEFPMNATKCDSESVDVETAPCSCQDCSIVCGFKPQPPPPPPAPM	300		
QY	271	---ALDSTFYLQMPGSLVLIILCSVFVVVTILL-----VGFRVAPARDKS	314		





```
Qy 1162 SYNVAISLNLVSAGMSVEFVSHITRSFAISTKPTWLERAKEATISMGSAFAGVAMTNL 1221
Db 1189 SLNVAISLNLVMSGSISEFVCSHITRAFTSTKGSVRSABEALAHMGSSVFSGITUTKF 1248
Qy 1222 PGLVLGLAKAQLQIIFFRNLNLITLLGLHLGLVFLPVLISVYGPDPVNP 1272
Db 1249 GGIVVLAFSAKQIEIFEFMYLAWLLGATHGLIFLPVLLSYIGPSVNA 1299

RESULT 4
US-09-462-136-6
; Sequence 6, Application US/09462136
; Patent No. 6426198
; GENERAL INFORMATION:
; APPLICANT: Carstee, et al.
; TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease
; FILE REFERENCE: 4239-53894
; CURRENT APPLICATION NUMBER: US/09/462,136
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: PCT/US98/13862
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/051,682
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1170
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-462-136-6

Query Match 19.2%; Score 1329; DB 2; Length 1170;
Best Local Similarity 28.8%; Pred. No. 3.4e-116;
Matches 373; Conservative 230; Mismatches 512; Indels 182; Gaps 46;

Qy 10 LLWAL-----LLRLAQSEPTTTHQPGYCAPYDCGKNP-----ELSGSLMTLSNVSLSN 60
Db 3 VLWIIALVGQLMRLVQG-----TATCAMYGNCGKKSFGNELPCFVPRSFEPVLS 54
Qy 61 TPARKITGDHLLILQKICPRLYTGPNTQACSAKQLVSLASLITKALLTRCPACSDNF 120
Db 55 ETSK-----LLVEVCGEEWKEVR-YACCTKDQVVALRDNLQAOPLISSCPACLNK 105
Qy 121 VNLHCHNTCSNQSLFNTRVAGOLGAGQLPAVVAEAFYQHSFAEQSYDCSRRVPAA 180
Db 106 NNLFCHTCARQGRFVNITKV-EKSKEDKDIAELDFVNMSSWASEFYDCKNKFSA 164
Qy 181 ATLAVGTCMGYGSALCNAQRWLNFGQDTGN--GLAPLDTIFHLLEPQAVGSGIQPLNE 238
Db 165 N---GYAMDLIGGAKNYSQFLAKFLGDAKPMLGSPFQIN KYDLANEE--KEWQEFND 218
Qy 239 GVARNESQGDVATCQDCAASCAPAIARPOALDSTFYLGQMPG---SLVLIILCSVP 295
Db 219 EVYACDDAQ----YKACACSDQCESCPHL-KP-LKDGVCVKGPLPCFSLVLIIFYTICALF 272
Qy 296 AVVTILVGRVAPARDKSKMVDPKKTSLSDK-----LSPSTHT-----LLGQFFQGW 345
Db 273 AFMYIICRKNKNGAMIVDDDIVESG--SLDESTNVFESFNNTETNFNGKLANLFTKVG 331
Qy 346 TWASWPLTILVLSVIPVALAAGLV-FTELTTDPVELWSPNQAARSEKAFHQHGF 404
Db 332 QFSVENPKLITTVFSIFVFSFIIFQVATLETDPINLWSKNSEKFEKEYFDNFGPF 391
Qy 405 FRTNQVILTAPNRSYYSDSL--LLGPNFSGILDLDLLELLELQELRLHLQWSPBAQ 462
Db 392 YRTEQIFVNVETGPNVLSYETLHWFDVENF-----ITBEL-----QSS 429
Qy 463 RNISLQIDICYAPLNPDTNLDYCINSLLQYFQNNRTLLLTANTQTLMGQTSQVDWKDH 522
Db 430 ENIGYQDLCFRP--TEDST-----CVIESFTQYFQ-----ALPNKDSWREL 470
Qy 523 LYCANAPLTFKDGITALALSCMADYGAPVPFPFLAIGYKGKDYSEAEALIMTFSLNYPAG 582
```

```
Db 471 QECGKFP-----VNCLPFTFQPKTNLL---PSDDDIILNAHAFVVTLLLTNH--- 514
Qy 583 DPLAQAKLWEBAFLBEMRAFQRRMAGMFQVTTAERSLEDEINRTTAEDLPFATSYIV 642
Db 515 ---TQSANRWEER-LBEYLLDLKVPGL-RISFNTEISLEKELNN--NDISTVAISYLM 567
Qy 643 IFLYISIALGSSYSWSRVMVDKATLGLGGVAVVLGAVMAAMGPFYSVLGIRSSIVILQV 702
Db 568 MFLYATWALRRKDGKTRLL-----LGISGLLIVLASIVCAAGFLTFLGLKSLIIIAEVI 621
Qy 703 PFLVLSVGADNIFIFVLEYQR-LPRRPGEPREVIHGRALGRVAPSMMLCSLSEACFLG 761
Db 622 PFLILAIGIDNIFLITHEYDRNCEQKPEYSIDQKIISAIGRMSPSILMSLLCOTGCEFLA 681
Qy 762 ALTPMPAVRTFALTSLGLAVILDFLLQMSAFVALLSDSKRQEARSLDVCCKRQELPPP 821
Db 682 AFVTMPAVHNFAIYSTVSIFNGVLQLTAYVTSLSLYEKRSNYKQIT----- 728
Qy 822 GQCEGLLLGPFQKAYAPFLHLHMTIRGVVLLFLALFGVSLYSYCHISVGLDQELALPKDS 881
Db 729 -GNEETKESFLKTFYFKMLTQ---KRLIIIIIFSAWFTSLVFLPEIOFGDQTLAVQDS 784
Qy 882 YLLDYFLFLNRYFEVGAPVYFVTTLGYNFSSEAGMNAICSS-AGCNNFSPQTKIQYATEF 940
Db 785 YLVDYFKDYSFLNVGPPVYVMVK-NLIDLTKRQKQKICGKFTTCERDSLANVLE---QE 840
Qy 941 PEQSYLAIPASSWVDDFIDWLTSS--CCELYISGPNKDKFCPSTVNSLNLKNCMSITM 998
Db 841 RHRSTITEPLANWLDYFMFLNPNQDCCRL---KKGTDEVCPSPFSRRRC-ETCFQ--Q 894
Qy 999 GSVR-----PSVEQFHKYLPWFNLDRPNIKCPKGLAAYSTSVNLTSDGQVLASRMAY 1052
Db 895 GSNVYNMSGPECKDMEYLSIWIN-APSDPCPLGGRAPYSTAL-VYNETSVSASVERTA 952
Qy 1053 HKPLKNSQDYTEALRAARELANITADLRKVPCTGPAPFVFPYITITNVFQYLTILPEG 1112
Db 953 HHPLRSQKDFIQAY-----SDGVRISSFPPELDMFAYSPPYIFFVQYQTLGFLT 1001
Qy 1113 LFMLSLCLVTPFAVSCLLGLDLRSGLLNLISVILVDTVGFPMALWDISYNAVSLNLIV 1172
Db 1002 LKLGSAIILIFFISSVFL-QNIRSSFLLALVVTMIIVDIGALMALIGISLNAVSLNLI 1060
Qy 1173 SAVGMSVEFVSHITRSFAI---STKPTWLERAKEATISMGSAFAGVAMTNLPGILVLGL 1229
Db 1061 ICVGLGVFCVHIVRSETVVPSETKKDANSRVLSYSLNTIGESVTKGITLTKFVGCVLAF 1120
Qy 1230 AKAQIQLIPFRNLNLITLLGLHLGLVFLPVLISYVG 1266
Db 1121 AQSKEFDVYFRMFWFTLIIVAALHALLFLPALLSLFG 1157
```

```
RESULT 5
US-09-462-136-9
; Sequence 9, Application US/09462136
; Patent No. 6426198
; GENERAL INFORMATION:
; APPLICANT: Carstee, et al.
; TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease
; FILE REFERENCE: 4239-53894
; CURRENT APPLICATION NUMBER: US/09/462,136
; CURRENT FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: PCT/US98/13862
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/051,682
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-462-136-9
```





```
Db 24 PCRAGGRRRTGGLRAAAPDRD-----YLHRPSYCDAAFALEQI--- 65
QY 285 LVLIILCSVFAVVTILLVGRVAPADKSKMVDPK-----KGTSLSDKLSFSTHTLL 337
Db 66 -----SKGATGRKAPLWLRAKFORLLFLGCGYIQNC 98
QY 338 GQFFQGWGTWASWPLTILVLSVIPVLAAGLVFTELTDPVELWSAPNSQARSEKAFH 397
Db 99 GRF-----LVUGLLIFGAFVGLKAANLETNVEELVVEVGGVRSELYNT 143
QY 398 DQHFQFFRTN-QVILTAPNRSSYRYSLLLPKNFSGILDLLELLE--LQERLRLH 454
Db 144 RQKICEAMFNQMLIQPKBEG-----ANVLITTEALLQHLDSALQASRVHV 190
QY 455 QWSPDQARNISLQICVAP-----LMPDNTSLYDCINSLLQVFNQNTLLLLTA 505
Db 191 YMYN-----ROWKLEHLCYKSGELITETGYMDQIIEYLPCLIIITPLDCFWEGAKLQSGTA 246
QY 506 NOTLMGO-----TSQVDWKDHFY-----CAN----- 527
Db 247 --YLLGKPLRTWTFDPLEFLEELKKINYQVDSWEEMLNKAEGVGHYMDRCLNPADPDC 304
QY 528 -----APL-----TFKOGT-----ALALSCMA 544
Db 305 PATAPNKNSKPLDMALVNGGCHGLSRKYMWBELIVGGTVKNSTGKLYSAHALQTMF 364
QY 545 DYGAIVFPFLATCGYKGYDSEAEALIMTFSLNNYPAGDPRLAQAOKLWEEAFLEBMAFQ 604
Db 365 QLMTPKQMY--EHPFGVEY-----VSHINWNE--DKAAAILLEANQRTYVEVHVOSV 411
QY 605 RMACGMFOVTFATERSLEDEINRTTAEDLPIFATSYIVIFLYISIALGYSYSSWSRVMD 664
Db 412 AQNSTQKVLSTT--TLDDILKFSFDSVIRVASGYLLMLAYACLTM--LRWD--CSKS 465
QY 665 KATILGGLGVAVLGAAMAGFFSYLGRSSILVILQVVPFLVLSVGADNIFIFVLEYQRL 724
Db 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGDDVFLAHAFSET 525
QY 725 PRPGEPREVIHGRALGRVAPSMCLCSLSEALCFPLGALTMPAVRTALTSLGLAVILDF 784
Db 526 GQNKRIPEDETRGECLEKRTGASVALTSISNVTAFPMALIPALRAFSLQAAVVVFN 585
QY 785 LQMSAFVALLSLDSKROEASRLDVCCC-----VKPQEL-----PPG 822
Db 586 AMVLLIFAILSMGLYREDRDLDFCCFTSPCVSRVIOVFQAYTDHNTVSPPPY 645
QY 823 QEGU----- 827
Db 646 SSHPAHETQITMQSTVOLRTEYDPHTHYTTAEPRESEISVQPVTVTQDTLSCQSPST 705
QY 828 -----LLGFQKAVAPFLHWTIRGVVLLFLALFGVSL 861
Db 706 SSTRDLLSQFSDSSHLCLBPPCTKWTLSFPAKHVAPFLPKAKVAVVIFLGLLGVSL 765
QY 862 YSMCHISVGLDQELALPKDSYLLDFLFLNRYFVGAPVYFVTTLGVNFSSEAGMNAICS 921
Db 766 YGTTVRDGLDLDIVPRETREYDFIAAQKYFSF-----YMW----- 803
QY 922 SAGCNFSFTQIKIYAT-----BPPQSYLAIPASS-----WVDDFIDWJ----- 961
Db 804 -----YIVTOKADYPNIQHLLYDLIHRFSFNKYVMLEENKQLPKMWLHYFRDMLQGLQD 857
QY 962 -----TPSS-----CCRLYISGPNKDKFCPSVTNSLNCNCKNSITWGS 1000
Db 858 AFDSWETGKIMPNYKNGSDGVLAYKLLVQTGSRDK--PIDISQLT--KQRLVDADGI 913
QY 1001 VRPSVEQFHXYL--PMFLND-----RPNKCPKGGLAAY--STSVMNLTSQDQ 1043
Db 914 INPSA--FYIYLTAWVSDNPVAYAAASQANIRPHRPWVHDK--ADYMPETRLRIPAAEP 968
QY 1044 VLASEFMAHYKPLKNSQDYTAALRAARELANITA-DLRKVPGTDPAPFVFPYITNVFY 1102
Db 969 IEYAQFPFLNGLRDTSDFVEAIEKVKRTICSNYTSLSGLSSYPNG-----YFF-----LFW 1018
```

---

```
QY 1103 EYUUTILPEGLFMLSCLVETFAVSCULLGLDIRSGLLNLLSIWMILVDTVGFMAWDIS 1162
Db 1019 EYUUTILPEGLFMLSCLVETFAVSCULLGLDIRSGLLNLLSIWMILVDTVGFMAWDIS 1077
QY 1163 YNAVSLNLVSAVMSVEFVSHITRSP--AISTKPTWLBRAKEATISMGSAAVFAVAMTN 1220
Db 1078 LSAVPVILLASVIGVIEFTVHVVALAFLTAIGDKN---RRAVLALEMFAVPLDQ-AVST 1133
QY 1221 LPGAIVLGLAKAOIQTIFFRMLNLLTLLGLHLHGLFVLPVLSYVG--PDVNPALALEQ- 1277
Db 1134 LLGVMLAGSEFDFIVRYFFAVLAILTILGVNLGLVLLPVLSPFGVPYEPVSPANGNLRL 1193
QY 1278 --KRAEAAVAVMVASCPNHPSPRVSTADNIYVNHVSFECSIKG 1317
Db 1194 PTPSPPPSVVRPAMPFGHTH--SGSDSDSEYSSQTTVSG 1233

RESULT 7
US-08-656-055-19
; Sequence 19, Application US/08656055
; Patent No. 6027882
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/656,055
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/540,406
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-656-055-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

QY 225 PGAVSGGIQPLNEGVARCNESQDDVATCSQDCAASCFAIARPQALDSTFVLGQMPGS 284
Db 24 PGRPAGGRRRTGGLRAAAPDRD-----YLHRPSYCDAAFALEQI--- 65
QY 285 LVLIILCSVFAVVTILLVGRVAPADKSKMVDPK-----KGTSLSDKLSFSTHTLL 337
```

Db 66 -----SKGATGRKAPLWLRRAKFORLLFKLGCYIQKNC 98  
QY 338 GQFFQGGTGWASWPLTILVSVIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAPH 397  
Db 99 GKF-----LVVGLLIFGAFAGVGLKAANLETNVELWVEVGRVSRRELYT 143  
QY 398 DQHFQGGTGWASWPLTILVSVIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAPH 1277  
Db 144 RQKIGSEAMFPMQIMQTPKEG-----ANVLTTEALLQHLDSALQASRVHV 190  
QY 455 QWSPQARNISLQDIQYAP-----LNPDTSLYDCCINSLLQYQNNRTLLLLTA 505  
Db 191 YMTN-----ROWKLEHCYKSGELITETGYMDQIIEYDPLCLITPLDCFWEGAKLOSQA 246  
QY 506 NOTLMGO-----TSQVMDKDHFLY-----CAN----- 527  
Db 247 --YLLGKPLPRTWNPDPLEBLELKKINQVDSWEEMLNKAEGVGHGMDRCPCLNPADPDC 304  
QY 528 -----APL-----TPKQGT-----ALALSQWA 544  
Db 305 PATAPNKNSTKPLDMALVNLGGCHGLSRKTMHQEBELIVGTVKSTGKLVSAHALQTMF 364  
QY 545 DYGAQVFPFLAIGYKGDYSEAEALIMTFLSNYPAGDPRLAQAQKLEWEEFLEEMRAQ 604  
Db 365 QLMTPKQMY---EHFAGYEY-----VSHINMNE---DKAAALEAMQRTYVVEVHQSV 411  
QY 605 RRMAGMFQVTTAERSLEDEINRTTAEDLPIFATSVIVIFLYISLALGSYSSMSRVMDVS 664  
Db 412 AQNSTQKVLSTFT--TWLDDILKFSFDSVIRVASGVLLMLAYACLTM---LRWD---CSKS 465  
QY 665 KATLGCGVAVVLGAVMAAGPFSYLGIRSSVLQVVPFLVSVGADNIFIFVLEBYORL 724  
Db 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAAATQVLPFLALGVGDVDFLLAHAFSET 525  
QY 725 PRPGPREVHIGRALGRVAPSMMLCSLSEAICFPLGALTTPAVTFTALTSGLAVILDF 784  
Db 526 GQNKRIPEDEBTGECUKRTGASVALTSISNVTAFMAALIPALRAFSLQAAVVVVF 585  
QY 785 LQMSAFVALLSDSKRQASRLDVCCC-----VKQOEL-----PPG 822  
Db 586 AMVLLIFPALLSMDLYREDRLDIFCCFTSPCVSRVIOVEQAYTHTDNTREYSPFPY 545  
QY 823 QSEGL----- 827  
Db 646 SSHSFAHQITMQSTVQLRTEYDHPHVVYTTAEPRSEISVQPVVTQDTLSCQSPST 705  
QY 828 -----LLGFFQKAYAPFLHMITRGVVLVLLFLALFQVSL 861  
Db 706 SSTRDLLSOFSSHLCLBPCTKTWTLSSFAEKHYAPFLPKAKVYVIFLFLGLVSL 765  
QY 862 YSMCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGPVYVVTTLGYNFSSEAGMNAICS 921  
Db 766 YGTRVRDGLDITDIVPRETREVDFIAAQKFSP-----YNM----- 803  
QY 922 SAGCNFFSTQKIQYAT-----EPPEOSYLAIPASS-----WVDDFDLWL----- 961  
Db 804 -----YIVTQADYPNIOHLLYDLHRSPSNKYVMLEENKQLPKMWLHYFRDQLQGLD 857  
QY 962 -----TPSS-----CCRLYISGPNKDKCPTVNSLNCNCSITMGS 1000  
Db 858 AFDSOWETKIMPNNYKNGSDGVLAYKLLVQTGSRDK--PIDISOLT--KORLVADAGI 913  
QY 1001 VRPSVEQPHKYL--PWFPLND-----RPNIKCPKGGLAAY--STSVNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAWNSNDPVAYASQANIRPHRPEWHDK--ADYMPETRLRIPAAP 968  
QY 1044 VLASRFMAYHKLKNSQDYTEALRAARELAANITA--DLRKKVFGTDPAPFVFFYITINVPY 1102  
Db 969 IEVAQPPFYNLGRDTSDFVEAIEKVRTICSNVTSLSGLSSYPNG-----YFP-----LFW 1018  
QY 1103 EQYLITPLPGLFMLSCLVPTFAVSCLLGLDLRSLGNLLSIVMLVDTVTFGMALWDIS 1162  
Db 1019 EQYIGLRHWWLLFISVVLACTFLVCAVFLNPNWTAGII--VMVLAALMTFELFGMGLIGIK 1077

QY 1163 YNAVSLINLVSAGMSVEFVSHITRSF--AISTKPTWLRKAETISMGSVAVFAGVAMTN 1220  
Db 1078 LSAVPVILLIASVIGVETVHVALAFLTAIGDKN---RRAVLALHEHMAPVLDG-AVST 1133  
QY 1221 LFGILVLGAKAQILQIFFRNLNLLITLGLHLGLVFLPVILSYVG--PDVNPALALEQ- 1277  
Db 1134 LLGLVLMAGSEDFIVRYFFAVLAAILTILGLVNLGLVLLPVLLSFFPGPYEVSANGINRL 1193  
QY 1278 --KRAEEAAVAVMVASCPNHPRSVSTADNIYVNHSPGSIKG 1317  
Db 1194 PTPSPPPPVVRFAFPMPGHTH--SGSDSDSEYSSQTTVSG 1233

RESULT 8

US-08-954-668-19  
; Sequence 19, Application US/08954668  
; Patent No. 6172200  
; GENERAL INFORMATION:  
; APPLICANT: SCOTT, MATHEW P  
; APPLICANT: GOODRICH, LISA V  
; APPLICANT: JOHNSON, RONALD L  
; TITLE OF INVENTION: Patched Genes and their Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley, Hoag & Eliot  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII(text)  
; CURRENT APPLICATION DATA:  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Vincent, Matthew P.  
; REGISTRATION NUMBER: 36709  
; REFERENCE/DOCKET NUMBER: SUV-003.06  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1447 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-954-668-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
QY 225 PQQAVSGGIQPLNEGVARCNESQGDVATCSQDCAASCPAIRPAQLDSTFYLQGMPS 284  
Db 24 PGRPAGGRRRTTGLRAAAPDRD-----YLRPSYCDAAAFALQI--- 65  
QY 285 LVLIILICSVFAVWTILLVGFVAPARDKSKMVDPK-----KGTSLSDKLSFSTHTLL 337  
Db 66 -----SKGATGRKAPLWLRRAKFORLLFKLGCYIQKNC 98  
QY 338 GQFFQGGTGWASWPLTILVSVIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAPH 397  
Db 99 GKF-----LVVGLLIFGAFAGVGLKAANLETNVELWVEVGRVSRRELYT 143  
QY 398 DQHFQGGTGWASWPLTILVSVIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAPH 1277

Db 144 RQKIGBEAMFNQMIQTPEKRG-----ANVLTTEALLQHLDSALQASRVHV 190  
QY 455 QVMSPEAQRNISLQDICYAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505  
Db 191 YMYN----RQWLEHLCKYSGBELITETGYMDQIIIEVLYPCLIIITPLDCFWEGAKLQSGTA 246  
QY 506 NQTLNGO-----TSQVDWKDHFY-----CAN----- 527  
Db 247 --YLLGKPLRWTFNDFPLEFLEBKINQVDSWEEMLNKARVGHGYMDRCLNPPADPDC 304  
QY 528 -----APL-----TFKOGT-----ALALSCMA 544  
Db 305 PATAPKNSTKPLDMALVNGGCHGLSRKYMHWQBELIVGTVKNSTGKLVSAHALQTMF 364  
QY 545 DYGAIPVFPFLAIGGYKGKDYSEAEALIMTFLSNLYPAGDPRLAQAOKLWEEAFLEEMRAFO 604  
Db 365 QLMTPKQMY---EHFKGVEY-----VSHINWNE-----DKAAAILLEAWQRTVVEVHQS 411  
QY 605 RRMAGMFQVTTAERSLEDEINRTTAEDLPFATSVIVIFLYISLALGSYSSWSRVMD 664  
Db 412 AQNSTQKVLSTFT--TTLDDILKSFSDSVIRVASGYLLMLAYACLTW---LRWD--CSKS 465  
QY 665 KATLGGVAVVYLGAVMAAMGFPSYLGIRSSVLIVQVPPFLVLSVGADNIFIFVLEYQRL 724  
Db 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFSET 525  
QY 725 PRPGEPREVIHGRALGRVAPSMMLCSLSEACFFLPGALTTPMPAVRTFALTSLGLAVILDF 784  
Db 526 QGNKRIPPEDRTGECLEKKTGASVALTSISNVTAFFMAALIPALRAFSLQAAVVVNF 585  
QY 785 LIQMSAFVALLSLDKROBASRLDVCCC-----VKPOEL-----PPG 822  
Db 586 ANVLLIFPALLSMDLYRREDRLDIFCCFTSPCSRVIQVEPQATVTHDNTRISPPPPY 645  
QY 823 QGEGSL----- 827  
Db 646 SSHSPAHEQTITMQSTQVLRTEYDPHTHYVTTAEPSEISQPVTVTQDTLSQSPBST 705  
QY 828 -----LLGFFQKAYAPELLHWITRGVULLFLALFGVSL 861  
Db 706 SSTRDLLQFSDSSLUHLEBPCTKWTLSFAEKHYAPFLPKAKVQVIFLGLGLVSL 765  
QY 862 YSMCHISYGLDQELALPKDSYLLDYFLFLNRYFEVGAQVYFVTTLGYNFSSEAGMNAICS 921  
Db 766 YGTRVRDGLDITDIVPRETREYDFIAAQKYFSF-----YNW----- 803  
QY 922 SAGCNFSFTQKIQYAT-----EFPQSYLAIPASS-----WVDDFIDWL----- 961  
Db 804 -----YIVTOKADYPNIQHLLYDLHRSFSNVKYVMLEENKQLPKWMLHYFRDMLQGLQD 857  
QY 962 -----TPSS-----CCRLXISGPNKDKFCPSVTNSLNCNKCMSITMGS 1000  
Db 858 AFDSWETGKMPNPNKNGSDGVLAYKLLVQTGSRDK--PIDISQLT--KORLVADAGI 913  
QY 1001 VRPSVEQFHKYL--PWFLLND-----RENKPCPKGLAAY--STSVALTSDGQ 1043  
Db 914 INPSA--FVIYLTAMVSNPDVAYAAQANIRPHPEWVHDK--ADYMPETRLAIPAEP 968  
QY 1044 VLASFPAYHKLKNSQDYTEALRAELAANITA-DLRKVPGTDPAPFEPYITNVFY 1102  
Db 969 IEYAQFPFLNGRLTDSFVEAIEKVRTICSNYTSLSGLSSYPNG-----YPF-----LFW 1018  
QY 1103 EOYLTPLEGFLMLSLCLVPTFAVSCLLGLDLRSGLNLLSIVMVLVDTVGFMAWDIS 1162  
Db 1019 EOYIGLRHWLLFISVVLACTFLVCVAFLLNPWTAGII--VMVLAJMTVELFGMGLGIK 1077  
QY 1163 YNAVSLINLVAGMSVEFVSHITRSP--AISTKPTWLERAKEATISMSGSAVFAGVAMTN 1220  
Db 1078 LSAPVWILLIASVGIGVEFTVHVALAFLTAIGDKN---RRAVLALEHMFAPVLDG-AVST 1133  
QY 1221 LPGAIVLGAQAQLIQIFFRNLNLLITLLGLHLGLVFLPVLISYVG--PDYNPALALEQ- 1277

Db 1134 LLGVMLAGSEDFIVRYFFFAVLAILTILGVNLGLVLLPVLLSFFPGYPVSPANGLNRL 1193  
QY 1278 --KRAEBAVAAVMVASCNHPSPRSVSTADNIYVNHVSFEGSIKG 1317  
Db 1194 PTPSPPPPSVVRPAMPFGHTH--SGSDSDSDSEYSSQTTVSG 1233  
RESULT 9  
US-09-268-140-5  
; Sequence 5, Application US/09268140  
; Patent NO. 6268176  
; GENERAL INFORMATION:  
; APPLICANT: Gemmill, Robert M.  
; APPLICANT: Drabkin, Harry A.  
; TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED  
; FILE REFERENCE: 93445-00004  
; CURRENT APPLICATION NUMBER: US/09/268,140  
; CURRENT FILING DATE: 2000-03-12  
; PRIOR APPLICATION NUMBER: US 60/077,723  
; PRIOR FILING DATE: 1998-03-12  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1447  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-140-5

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 2460; Indels 421; Gaps 47;

QY 225 PGQAVSGIOPNEGVARNESQDDVATCSCQDCAACPAIARPOALDSTFFYLQMPGS 284  
Db 24 PGRPAGGRRRTGGLRRAAAPDRD-----YLRHPSYCDAAFALEQI--- 65  
QY 285 LVLLIILCSVFAVVTTILLVGRVAPARDKSKWDPK-----KGTSLSKLSFSTHTLL 337  
Db 66 -----SKGATGRKAPLWRAKFORLLFLKGCYIQKNC 98  
QY 338 GQFQFGGTWVSWNPLTILVSVIPVVALAAGLVFTLTDTPVELMSAPNSQARSEKAFH 397  
Db 99 GKF-----LVVGLLIIFGAFVAGLKAANLETNVEELWVEVGGVRSRELYNT 143  
QY 398 DQHFPGPPFTN-QVILTPAPNRSSRYVDSLGLPKNFGSGLDLDLLLELLE--LQERLRL 454  
Db 144 RQKIGBEAMFNQMIQTPEKRG-----ANVLTTEALLQHLDSALQASRVHV 190  
QY 455 QVMSPEAQRNISLQDICYAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505  
Db 191 YMYN----RQWLEHLCKYSGBELITETGYMDQIIIEVLYPCLIIITPLDCFWEGAKLQSGTA 246  
QY 506 NQTLNGO-----TSQVDWKDHFY-----CAN----- 527  
Db 247 --YLLGKPLRWTFNDFPLEFLEBKINQVDSWEEMLNKARVGHGYMDRCLNPPADPDC 304  
QY 528 -----APL-----TFKOGT-----ALALSCMA 544  
Db 305 PATAPKNSTKPLDMALVNGGCHGLSRKYMHWQBELIVGTVKNSTGKLVSAHALQTMF 364  
QY 545 DYGAIPVFPFLAIGGYKGKDYSEAEALIMTFLSNLYPAGDPRLAQAOKLWEEAFLEEMRAFO 604  
Db 365 QLMTPKQMY---EHFKGVEY-----VSHINWNE-----DKAAAILLEAWQRTVVEVHQS 411  
QY 605 RRMAGMFQVTTAERSLEDEINRTTAEDLPFATSVIVIFLYISLALGSYSSWSRVMD 664  
Db 412 AQNSTQKVLSTFT--TTLDDILKSFSDSVIRVASGYLLMLAYACLTW---LRWD--CSKS 465  
QY 665 KATLGGVAVVYLGAVMAAMGFPSYLGIRSSVLIVQVPPFLVLSVGADNIFIFVLEYQRL 724  
Db 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFSET 525  
QY 725 PRPGEPREVIHGRALGRVAPSMMLCSLSEACFFLPGALTTPMPAVRTFALTSLGLAVILDF 784

Db 526 GQKRIPFEDTGECLRTGASVALTSISNVTAFMAALIPFALRAFSLOAAVVFNF 585  
Qy 785 LLQMSAFVALLSDSKQASRLDVCC-----VKQBEL-----PPPG 822  
Db 586 AMVLLIPFALLSMDLYRERDRLDIFCCFTSPCVSRVQVEPQAYTTHDNTYSPPPY 645  
Qy 823 QSGGL-----827  
Db 646 SSHSFAHETQITMQSTVQLRTEYDPHTHYVYTTAEPRSEISVQPVVTQTLSQCSPEST 705  
Qy 828 -----LLGPFQKAYAPFLHWTIRGVLLHFLFALFGVSL 861  
Db 706 SSTRDLLSQSDSLHCLPECTKWTLSFAEKHYAPFLKPKAKVWVIFLFLGLLGVSL 765  
Qy 862 YSMCHISVGLDQELALPKDSYLDYFLFLNRYEFGVAPVYVTTLLGNFSESSEAGMAICS 921  
Db 766 YGTRVRDGLDLDIVPRETREDYFTAAQPKYFSF-----YNM-----803  
Qy 922 SAGCNFPSTQKIQYAT-----EPPEQSYLAIPASS-----WVDDFDWL-----961  
Db 804 -----YIVTQKADYENIQHLLYDLHRSFNSVNVKVMLEENKQLPKWMLHYFRDWLQGLQD 857  
Qy 962 -----TPSS-----CCRLYISGNKDKCPCSTVNSINCLKNCSITMGS 1000  
Db 858 AFSDWETGKIMPNNYKNGSDDGVLAYKLLVQTGSRDK--PIDISQLT--KQRLVDADGI 913  
Qy 1001 VRPSVQFHKYL--PWFLND-----RPNIKCPKGLAAY--STSVNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAWNSNDPVAASAQANIRPHRPWEVHDK---ADYMPETRLURIPAAEP 968  
Qy 1044 VLASRFMAXHKLKNSQDYTEALRAARELANITA--DLRKVPQGTDPAFEPVFTYITNVFY 1102  
Db 969 IEVAQPFYFLNGLRDTSDFVEATEKVRTICSNVTSGLSSYPNG-----YPP-----LFW 1018  
Qy 1103 EQVLTTLPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSVMLVLTDTVGFMALWDIS 1162  
Db 1019 EQYIGLRHLLFISVTLACTFLVCVAVFLNPNWTAGII--VMVLALMTVLEFGMGLIGIK 1077  
Qy 1163 YNAVSLINLVSAGMGSVEFVSHITRSF--AISTKPTWLBRAKEATISMGSAVPAGVAMTN 1220  
Db 1078 LSAPVVLIIASGVIGVETVHVHVAFLTAIGDKN---RRVLALEHMFAPVLDG-AVST 1133  
Qy 1221 LFGILVLGAKAQLIQIPFRLMLLITLLGLHGLVFLVILSYVG--PDVNPALALEQ- 1277  
Db 1134 LLGVLMAGSEDFIVRYFFAVLAILTLGLVNLGLVLLPVLISFFGYPYEVSPANGINRL 1193  
Qy 1278 --KBAEAVAAVWVASCNHPSPRVSTADNIYNHSEGSIGK 1317  
Db 1194 PTPSPPPPSVVRFAFMPGHTH--SGSDSDSEYSSQTTVSG 1233

RESULT 10

US-08-918-658-19  
; Sequence 19, Application US/08918658  
; Patent No. 6429354  
; GENERAL INFORMATION:  
; APPLICANT: SCOTT, MATHEW P  
; GOODRICH, LISA V  
; JOHNSON, RONALD L  
; TITLE OF INVENTION: Patched Genes and their Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/918,658  
FILING DATE: 22-Aug-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/656,055  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/540,406  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: a60190-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-08-918-658-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
Qy 225 PQQAVSGSIQPLNEGVARCNESQGDVATCSQDCAASCPAIPARQALDSTFVLGQMPGS 284  
Db 24 PCRAGGRRRTTGLRRAAAPDRD-----YLHRPSYCDAAFALEQI--- 65  
Qy 285 LVLIILCSVFAVITLLVGFVAPARDKSKWDPK-----KGTSLSDKLSFSHTLL 337  
Db 66 -----SKGKATGRKAPLWRAKFORLLFKLGCYIQQNC 98  
Qy 338 GQFFQGWGTWASWPLTILVSVIPVVAAGVFTLTDPVELWSAPNSQARSEKAFH 397  
Db 99 GKPF-----LVGGLIFGAFVAGLKAANLETNVEELMVEVGVRSRELYT 143  
Qy 398 DQHFQFPFRTN-QVILTAPNRSSRYDSSLGPKNPSGILDLDLLELLE--LQERLRHL 454  
Db 144 RQKIGBEAMFNQPMIQPKESG-----ANVLTTEALLQHLDSALQASRVHV 190  
Qy 455 QVWSPEAQRNLSQDICVAP-----LNPDNTSLYDCINSLLQYFQNNRTLLLLTA 505  
Db 191 YMYN----RQWKLEHLCYKSGELITETGYMDQIIIEYLYPCLITPLDPCFWEKALQSGTA 246  
Qy 506 NQTLMGQ-----TSQVDMKDHPLY-----CAN----- 527  
Db 247 --YLLGKPLRWTNFDPLEFLEELKKNYQVDSWEEMLNKAIEVGHYMDRCPINPADPC 304  
Qy 528 -----APL-----TFKDG-----ALALSCMA 544  
Db 305 PATAPKNKSTKPLDMALVNLGCGHGLSKRYMHWQBELIVGTVTKSTGLVSAHALQTFW 364  
Qy 545 DYGAVPFPFLAIGGYKDYSEAEALIMTFSLNPNYPAGDPRLQAQAKMEEAELEENRAFO 604  
Db 365 QLMTPEKQMY---EHFKGYEY-----VSHINWNE---DKAAAILLEAWQRTYVVEVHVQSV 411  
Qy 605 REMAGMFQVTTFAERSLEDEINRTTAEDLPFATSVYIVFLYISALGYSYSSWSRVWVDS 664  
Db 412 AQNSTQKVLSPFT--TTLDDILKSFSDSVIRVASGYLLMLAYACLTM---LRWD--CSKS 465  
Qy 665 KATLGLGVAVVLGAVMAAMGFFSYLGIRSSSILVILQVVPFLVLSVADNIFIFVLEYQRL 724  
Db 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAATQVLPFLALGVGVDDVDFLLAHAFST 525  
Qy 725 PRPQGEPEVHIGRALGRVAPSMMLCSLSEACFPFGALTPMPAVRTTALTSGLAVIDLF 784

526 GONKRIPEPTDRTGECLKRTKRTGASVALTISNVTAFMAALIPALRAFSLQAAVVVNF 585  
785 LLOWSAFVALLSLDSKROEASRLDVCC-----VKPOEL-----PPG 822  
586 AMVLLIFPAILSMDLYREDRLDIFCCFTSPCSRVIQVPEQAYTDTHTDTRYSPPPY 645  
823 QGEGL----- 827  
646 SSHSFAHETQITMSTVQLRTEYDPHTHYVTTAEPRSEISVQPVTVTQDTLSQSPST 705  
828 -----LLGPFQKAYAPFLHWTIRGVVLLFLFLAFVGS 861  
706 SSTRLLSQFSSSLHCLPEPCTKTWLTSSFAEKHYAPFLPKAKVVVIFLGLLGS 765  
862 YSMCHISVGLQELALPKDSVLLDYFLFLNRYFEVGAPVYFVTILGYNFSSEAGNAICS 921  
766 YGTRVRDGLDITDVPRETREYDFIAAQKYFSF-----YNN----- 803  
922 SAGCNFSTKIOYAT-----EFPEQSYLAI PASS-----WYDDFDWL----- 961  
804 -----YIVTQADYPNIQHLLYDLHRSFNVKVMLEENKQLPKWMLHYFRDMLQGLD 857  
962 -----TPSS-----CRLYISGNKKKPCPSTVNSLNCNCKMSITWGS 1000  
858 AFDSWETGKIMPNNYKNGSDGVLAYKLVQTGSRDK--PIDISQLT--KQRLVDADGI 913  
1001 VRPSVEQHKYL--PWFLND-----RPNKCPKGLAAY--STSVMLTSDGQ 1043  
914 INPSA--FYIYLTAWVNDPWAYAASQANIRPHRPWWHDK---ADYMPETRLRIPAAEP 968  
1044 VLASRFMAHYKPLKNSQDYTEALRAARELANITA-DLRKVPGTDPAPVFPYITNVFY 1102  
969 IEYAQFPYNGLRDTSDFVEAIEKVRTICSNYISLGLSSYPNG-----YFP-----LFW 1018  
1103 EOYLTLPEGFLMSLCLVPTFAVSCLLLGLDLSGLNLSIVMILVDTVGFMAWDIS 1162  
1019 EOYIGLRHLLIFISVWLACTFLVCAPFLNPWTAGII-VVVALMTVELFGMMGLIGIK 1077  
1163 YNAVSLINLVAGNSVEFVSHITSEF--AISTKPTWLERAKEATISNGSAVFAGVAMTN 1220  
1078 LSAPVSVILIASVGIGVEFTVHVVALAFLTAIGDKN---RRAVLALEHFAFVLDG-AVST 1133  
1221 LPGAIVLGLAKAQITQIFFFRLNLTLLGLLHGLVFLPVLISYVG--PDVNPALALEQ- 1277  
1134 LLGVLMLAGSEFDFIVRYFAVLAITLILGLVNLGLVLLPVLSPFGPYPEVSPANGLNL 1193  
1278 --KRAEEAAMVAVASCPNHSRSTADNIYVNHSEFGSING 1317  
1194 PTPSPPPPVVRFAFPPGHTH--SGSDSDSEYSSQTTVSG 1233

RESULT 11

US-09-724-631-19  
; Sequence 19, Application US/09724631  
; Patent No. 6551782

GENERAL INFORMATION:

APPLICANT: SCOTT, MATHEW P  
; GOODRICH, LISA V  
; JOHNSON, RONALD L  
; TITLE OF INVENTION: Patched Genes and their Use  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: US  
; ZIP: 94111

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/724-631  
; FILING DATE: 28-No. 6551782-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/656,055  
; FILING DATE: 1996-05-31  
; APPLICATION NUMBER: 08/540,406  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rowland, Bertram I  
; REGISTRATION NUMBER: 20015  
; REFERENCE/DOCKET NUMBER: a60190-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-781-1989  
; TELEFAX: 415-398-3249  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1447 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
; US-09-724-631-19  
; Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
; Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
; Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
; 225 PGQAVSGGIQPLNEGVARCNESQGDVATCSCODCAASCPAIPARQALDSTFYLGQMPS 284  
; 24 PGRPAGGRRRTTGLRRRAAPDRD-----YLRPSYCDAAFALEQI--- 65  
; 285 LVLIILCSVFVVVITLLVGRVAPARDKSKVDPK-----KGTSLSDKLSFSTHLL 337  
; 66 -----SKGKATGRKAPLMWRKAFQRLFLKLCGYIQKNC 98  
; 338 GQPFQGWGTWVSWPLTILVLSVIPVVALAAGLVFTLTDPVELASAPNSQARSEKAFH 397  
; 99 GK- -----LVVGLLIIFGAFVAGUKAANLETNVBELVVEVGGVRSRELYNT 143  
; 398 DQHGFPFRTN-QVILTAPNRSSRYVDSLGLPKNFSGIILDLILLLELLE--LQERLRLH 454  
; 144 RQKIGEEAMNPQLMIQTPKEG-----ANVLTTEALLQHLDSALQASRVHV 190  
; 455 QVWSPEAQRNLSIDICYAP-----LNPNSTSLYDCCINSLLQYFQNNRTLLLLTA 505  
; 191 YMTN-----RQWKLEHLCKYSGELITETGYMDQIIEYLYPCLIIITPLDCFWEGAKLQSGTA 246  
; 506 NQTLMGO-----TSQVDMKDHFLY-----CAN----- 527  
; 247 --YLLGKPPLRWTNFDPLEBELKKINYQVDSWEEMLNKAKEVGHGYMDRPLNADPDC 304  
; 528 -----APL-----TFKQGT-----ALALSCMA 544  
; 305 PATAPNKNSTKPLDMALVNLGCGCHLSRKYMHQEEELIVGCTVKNSTGKLVSAHALQTMF 364  
; 545 DYGAVPFPFLAIGYKGYSEBALIMTSLNNYPAGDPRLAQAKLWEAFLEEMRAFQ 604  
; 365 QLMTPKQMY---BHFKGYEY-----VSHINWNE-----DKAAAILAEAWQRTYVEVHQSV 411  
; 605 RRMAGMFQVFTTAERSLEDEINRTABDLPIFATSYIVIFLYISLALGYSYSSRSRVNVD 664  
; 412 AQNSTQKVLSTFT--TLLDDILKSFDSVSVIRVAGYLLMLAYACLTM---LRWD--CSKS 465  
; 665 KATLGLGGVAVLGVAMAMGFFSYLGRSSVLVQVVPFLVSVGADNIFIFVLEYQRL 724  
; 466 QGAVGLAGVLLVALSVAAGLGLCLIGISFNAATTQVLPFLALGVGVDDVFLLAHAFSET 525  
; 725 PRPGEPREVHIGEALGRVAPSMLLCSLSAI CFFLGLATPMPAVRTFALTSLGLAVILDF 784  
; 526 GONKRIPEPTDRTGECLKRTKRTGASVALTISNVTAFMAALIPALRAFSLQAAVVVNF 585



Db 646 SSHFAHETQITMSTQVLRTEYDPHTHYVYTTAEPRSEISVQPVVTQDTLSQCSPEST 705  
Qy 828 -----LLGFFQKAYAPFLHWHITRGVWVLLFLALFGVSL 861  
Db 706 SSTRDLSQFSSSLHCLPECTKTWTLSSFAEKHYAPFLPKAKVWVIFLFLGLLGVSL 765  
Qy 862 YSMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGAPVYFVTTGLNFSSEAGMNAICS 921  
Db 766 YGTRVRDGLDLDIVPRETREYDFIAAQKFYFSF-----YNN----- 803  
Qy 922 SAGCNFSFTQKIYAT-----EFPEQSYLAIPASS-----WVDFIDWL----- 961  
Db 804 -----YIVTQKADYPNIQHLLYDLHRSFNSVYVMLEENKQLPKWMLHYFRDMLQGLQD 857  
Qy 962 -----TPSS-----CCRLYISGNKDKFCPSTVNSLNCNKCMSITWGS 1000  
Db 858 AFSDWETGKIMPNNYKNGSDGVLAYKLLVQTGSRDK--PIDISQLT--KQRLVDADGI 913  
Qy 1001 VRPSVEQPHKYL--PWFLND-----RPNIKCPKGGGLAAY--STSNNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAWVNSDPVAYAASQANIRPHRPWVHDK--ADYMPETRLRIPAAEP 968  
Qy 1044 VLASRFWAYHKPLKNSQDYTEALRAARELANITA--DLRKVPGTDPAPFVPPYITITNVFY 1102  
Db 969 IEYAQFPFYNGLRDTSDFEAEIKVRTICSNYTSLSGLSSYPNG-----YPP-----LFW 1018  
Qy 1103 EGYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSIWMLVDTVGFMALWDIS 1162  
Db 1019 EGYTLGRHLLLFISVVLACTFLVCVAFLLNPWTAGII--VMVLAMTVLFGMGLIGIK 1077  
Qy 1163 YNAVSLINLVSAGVMSVEFVSHITRSF--AISTKPTWLERAKEATISMGSAVFAGVAMTN 1220  
Db 1078 LSAVPVILLASVIGIEFTVHVALAFETALCDKN--RRAVLALEHMFAPVLDG-AVST 1133  
Qy 1221 LPGAIVLGLAKAQLIQIHPFRANLITLLGLHLGLVFLPVLISYVG--PDVNPALALEQ- 1277  
Db 1134 LGLVLMAGSEDFIVRYFFAVLAILITLGLVNLGLVLLPVLLSFFGPPVPSVPSANGLNRL 1193  
Qy 1278 --KRAEAAVAAVMVASCNHRSPVSTADNIYVNHISFEGSING 1317  
Db 1194 FTPSPPEPPSVRFAMPFGHTH--SGSDSDSDSEYSSQTTVSG 1233

## RESULT 13

US-09-807-007-6  
; Sequence 6, Application US/09807007  
; Patent No. 6981833  
; GENERAL INFORMATION:  
; APPLICANT: ZAPHIROPOULOS, Peter et al.  
; TITLE OF INVENTION: A NOVEL COMPONENT IN THE HEDGEHOG SIGNALLING PATHWAY  
; CURRENT APPLICATION NUMBER: US/09/807,007  
; CURRENT FILING DATE: 2001-04-06  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 1447  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-807-007-6

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
Qy 225 PGAVGSGIOPLNEGVARCNESQGDVATCSQDCAASCPAIRPQALDSTFYLGQMPGS 284  
Db 24 FGRPAGGRRRTTGLRRAAPDRD-----YLHRPSYCDAAFALEQI--- 65  
Qy 285 LVLIILICSVAVNTILLVGFVAPARDKSNVDPK-----KGTSLSDKLSFSTHLL 337  
Db 66 -----SKGATGRKAPLWLRRAKFORLLFLKGLCYIQKNC 98

Qy 338 GOFFQCGMTWVASWPLTILVLSVIPVALAAGLVFTTELTTDPVELMSAPNSQARSEKAFH 397  
Db 99 GKF-----LVVGLLIIFGAFVAGUKAANLETNVELWVEVCGVRSRELNYT 143  
Qy 398 DQHFQGFPPFTN-QVILTAPNRSRYDSLLGPKNFSGILDLDLLELLE--LQERLRLH 454  
Db 144 RQKIGEEAMFNPQLMIQTPEEG-----ANVLTTALLQHLQALQASRVHV 190  
Qy 455 QVWSPQAQRNISLQDICYAP-----LNPNTSLYDCINSLLQYFQNNRTLLLLTA 505  
Db 191 YMYN-----ROWKLEHLCYKSGELITETGYMDQIIEYLYPCLIIITPLDCFWEGAKLQSGTA 246  
Qy 506 NQTLMGQ-----TSQVDWKDHFY-----CAN----- 527  
Db 247 --YLLGKPLRWTNFPDPLEBELKKNYQVDSWEEMLNKAEVGHGYMDRPLNPDAPDC 304  
Qy 528 -----APL-----TFKQGT-----ALALSCMA 544  
Db 305 PATAPNKNSTKPLDMALVNLGCHGLSRKYMHWQEEILVGGTVKNSTGKLVSAHALQTMF 364  
Qy 545 DYCAPVFPFLAIGGYKGDYSEAEALIMTFLSNYPAGDPRLAQAKLWEAEFLMEAFQ 604  
Db 365 QLMTPKQMY---BHFPGYGY-----VSHINWNE-----DKAAAILAEAWQTYVEVHQS 411  
Qy 605 RRMAGMFQVFTTAERSLEDEINRTTAEDLPFATSYIVIFLYISLALGSYSSRSRVNVD 664  
Db 412 AQNSTQKVSFTT--TILDDILKSFSDSVIRVAGVYLLMLAYACLTM--LRWD--CSKS 465  
Qy 665 KATLGLGVAVVLGAVMAAMGFFSYIGIRSSLVILQVPPFLVLSVSGADNIFIFVLEYQRL 724  
Db 466 QCAVGLAGVLLVALSVAAGLGLCSLIGISFNAATQVLPFLALGVGVDDVFLLAHAFSET 525  
Qy 725 PRPGPREVHIGRAGRVAPSMLLCSLSEAI CFELGALTMPBNAVRTFALTSLGLVILDF 784  
Db 526 GQNKRIFFEDRTGECRLKRTGASVALTSISNVTAFMAALIPIPALRAFSLQAAVVVVFNF 585  
Qy 785 LLOMSAFVALLSDSKQEAERLDVCCC-----VKPOEL-----PPPG 822  
Db 586 AMVLLIFPALLSMDLTYREDRRLDIFCCFTSPCVSRVIOVEPOAYTDTHTNRYSPPPY 645  
Qy 823 QGBGL----- 827  
Db 646 SSHFAHETQITMSTQVLRTEYDPHTHYVYTTAEPRSEISVQPVVTQDTLSQCSPEST 705  
Qy 828 -----LLGFFQKAYAPFLHWHITRGVWVLLFLALFGVSL 861  
Db 706 SSTRDLSQFSSSLHCLPECTKTWTLSSFAEKHYAPFLPKAKVWVIFLFLGLLGVSL 765  
Qy 862 YSMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGAPVYFVTTGLNFSSEAGMNAICS 921  
Db 766 YGTRVRDGLDLDIVPRETREYDFIAAQKFYFSF-----YNN----- 803  
Qy 922 SAGCNFSFTQKIYAT-----EFPEQSYLAIPASS-----WVDFIDWL----- 961  
Db 804 -----YIVTQKADYPNIQHLLYDLHRSFNSVYVMLEENKQLPKWMLHYFRDMLQGLQD 857  
Qy 962 -----TPSS-----CCRLYISGNKDKFCPSTVNSLNCNKCMSITWGS 1000  
Db 858 AFSDWETGKIMPNNYKNGSDGVLAYKLLVQTGSRDK--PIDISQLT--KQRLVDADGI 913  
Qy 1001 VRPSVEQPHKYL--PWFLND-----RPNIKCPKGGGLAAY--STSNNLTSDGQ 1043  
Db 914 INPSA--FYIYLTAWVNSDPVAYAASQANIRPHRPWVHDK--ADYMPETRLRIPAAEP 968  
Qy 1044 VLASRFWAYHKPLKNSQDYTEALRAARELANITA--DLRKVPGTDPAPFVPPYITITNVFY 1102  
Db 969 IEYAQFPFYNGLRDTSDFEAEIKVRTICSNYTSLSGLSSYPNG-----YPP-----LFW 1018  
Qy 1103 EGYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSIWMLVDTVGFMALWDIS 1162  
Db 1019 EGYTLGRHLLLFISVVLACTFLVCVAFLLNPWTAGII--VMVLAMTVLFGMGLIGIK 1077  
Qy 1163 YNAVSLINLVSAGVMSVEFVSHITRSF--AISTKPTWLERAKEATISMGSAVFAGVAMTN 1220



1078 LSAPVVLIASVGIGVEFTVHVALAFLTAIGDKN---RRVLALEHMFAPVLDG-AVST 1133  
1221 LPIGLVLGAKAQLIQIPFPRLLMLLTLLGLLHGLVPLVILSYVG--PDVNPALALEQ- 1277  
1134 LLGVLMLAGSEPDFIVRYFAVLAIITILGVNLGLVLLPVLISFFGYPYEVSPANGLNRL 1193  
1278 --KRAEEAANAAMVASCNPHSPSRVSTADNIYVNHSPFGSIKG 1317  
1194 PTPSPPPPPSVRFAMPFGHTH--SGSDSDSEYSQTTVSG 1233

RESULT 14

US-09-754-032-19  
Sequence 19, Application US/09754032  
Patent No. 6921646  
GENERAL INFORMATION:  
APPLICANT: SCOTT, MATHEW P  
GOODRICH, LISA V  
JOHNSON, RONALD L  
TITLE OF INVENTION: Patched Genes and their Use  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: US  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/754,032  
FILING DATE: 03-Jan-2001  
CLASSIFICATION DATA:  
PRIORITY APPLICATION DATA:  
FILING DATE: 06-Oct-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rowland, Bertram I  
REGISTRATION NUMBER: 20015  
REFERENCE/DOCKET NUMBER: a60190-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1447 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-754-032-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;  
Best Local Similarity 22.0%; Pred. No. 1.1e-51;  
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;  
QY 225 PQOAVSGIQPLNEGVARCNESQGDVATCSQDCAASCFATARPQALDSTPYLGMPGS 284  
DB 24 PGRPAGGRRRTTGLRRRAAPDRD-----YLHRSYCDAAAFALSGI--- 65  
QY 285 LVLIILICSVFAVWTVLLVGFVAPARDKSRMVDPK-----KGTSLSDKLSFSFTHLL 337  
DB 66 -----SKGATGRKAPLWLRAKFORLLFLKGCYIQKNC 98  
QY 338 GQFFQGWGTWASWELTILVLSVIPVVALAAGLVFTLTDPVELWSPNPSQARSEKAPH 397  
DB 99 GRF-----LVVGLLIFGAPVGLKAANLETNVEELWVEVGGVRSRELNYT 143

QY 398 DQHFQFFRTN-QVILTAPNRSSYRYSLLGPKNPSGILDLDLLELLE--LQERLRLH 454  
DB 144 RQKIGBEAMFNQLMIQTPEKG-----ANVLTTEALLQHLDSALOASRVHV 190  
QY 455 QWSPSPAQRNISLQDICYAP-----LNPDTSLYDCCINSLSLQYQNNRTLLLLTA 505  
DB 191 YMYN-----RQWKLEHLCYKSGELITBTGYMDQIIEYLYPCLIITPLDCWEGAKLOGTA 246  
QY 506 NOTLMGO-----TSQVDWKDHFY-----CAN----- 527  
DB 247 --YLLGKPLRWTFNDFLEBLELKINYQVDSWEEMLNKAEVGHGYMDRCPCLNPDAPDC 304  
QY 528 -----APL-----TFKQGT-----ALALSMA 544  
DB 305 PATAPNKNSTKPLDMALVNLGGCHGLSRKYMHEQBELIVGGTVKNSTGKLSAHALQTMF 364  
QY 545 DYGAVPFPFLAIGGYKGDYSEABALIMTFSLNNYPAGDPRLAQAKLWEAEFLERNRQF 604  
DB 365 QLMTPKQMY---EHFKGYEY-----VSHINNE---DKAAAILEAWQRTYVVEVHQSV 411  
QY 605 RRMAGMFQVTTAERSLEDEINRTTAEDLPIFATSYVIVFLYISLALGSYSSKSRVWVDS 664  
DB 412 AQNSTQKVLSFTT-TLDDILKSPDSVIRVASGYLLMLAYACLTM---LRWD---CSKS 465  
QY 665 KATLGLGGVAVVVLGAVMAAMGFFSYLIGIRSSLVILQVVPFLVLSVGADNIFVLEYQRL 724  
DB 466 QGAVGLAGVLLVALSVAAGLGLCSLIGISFNAAQTQVLPFLALGVGDVDFLLAHAFSET 525  
QY 725 PRPGEPREVIHGRALGRVAPSMMLCSLSEACFFLPGALTMPAVRTFALTSLGLAVILDF 784  
DB 526 GQNKRIPEPDRTGECLEKRTGASVALTSISNVTAFMAALIPALRAFSLQAQVAVVVF 585  
QY 785 LIQMSAFALLSLDSKROEASRLDVCCC-----VKQEL-----PPPG 822  
DB 586 AMVLLIFPAILSMWLYRREDRLDIFCCFTSPCVSRVIOVPEQAYTDTHTNRYSPPPY 645  
QY 823 QGEGE----- 827  
DB 646 SSHSFAHETQITMQSTVQLRTEYDPTHVYVTTASPRSEISVQPVTVTQDTILSCQSPST 705  
QY 828 -----LLGFPKQAYAPFLHWHITRGVLLLFLALFGVSL 861  
DB 706 SSTRDLLSQSDSSLHLCLEPPCTKWTLSFAEKHYAPFLKPKAKVWVFLFLGLIGVSL 765  
QY 862 YSMCHISVGLQDELALPKDSYLLDYFLNRYFEVGAPVYFVTILGYNFSSEAGWNAICS 921  
DB 766 YGTTVRDGLDLTDIVPRETREYDFIAAQKYFSF-----YNN----- 803  
QY 922 SAGCNNFSFTQKIQYAT-----EPPEQSYLAIPASS-----WVDDDFIDWL----- 961  
DB 804 -----YIVTQKADYPNIQHLLYDLHRSFNSVYKVMLEENKQLPKMWLHYFRDMLQGLQD 857  
QY 962 -----TPSS-----CCRLYISGPNKDKPCPSTVNSLCKNCSITMGS 1000  
DB 858 AFDSWETGKIMPNNYKNGSDGVLAYKLLVGTGSRDK--PIDISQLT--KQRLVADAGI 913  
QY 1001 VRPSVEQTHKYL-PWFLND-----RNIKCPKGLLAAY--STSVNLTSQDQ 1043  
DB 914 INPSA--FYIYLTAWVSNPDVAYAAASQANIRHPRPEWVHDK---ADYMPETRLRIPAAEP 968  
QY 1044 VLASRFMAHYKPLKNSQDYTEALRAARELANITA-DLRKVPGTDPAPFVFPYTTITNVPY 1102  
DB 969 IEYAGFPFYNGLRDTSDFVEAEIKVRTICSNYSLSGLSSYPNG-----YPF-----LFW 1018  
QY 1103 EGYLTILPEGLFMASLCIVPTFAVSCLLGLDLRSLGLNLLSIVMLVDTVGFMAWDIS 1162  
DB 1019 EGYIGLRHWLLIFTISVWLACTFLVCVAFLLNPWTAGII-VVMALMTVELFGMGLIGIK 1077  
QY 1163 YNAVSLINLVSAVGSVEFVSHITRSP--AISTKPTWLERAKEATISMGSAVPAGVAMTN 1220  
DB 1078 LSAPVVLIASVGIGVEFTVHVALAFLTAIGDKN---RRVLALEHMFAPVLDG-AVST 1133

```
Qy 1221 LPGAIVGLAKAQLQIIFERLNLITLLGLHGLVFLPVLSYVG--PDVNPALALEQ- 1277
Db 1134 LLGVMLAGSEDFIVRVFFAVLAILTILGLVGLVLLPVLSFGPYPEVSPANGLNRL 1193
Qy 1278 --KRAEEAAVAAVMVASCNHPSPSVSTADNIYVNHSEFGSIKG 1317
Db 1194 PTPSPPEPPSVVRFAAMPQGH--SGSDSDSEYSSQTIVSG 1233

RESULT 15
US-08-916-140-19
; Sequence 19, Application US/08916140
; Patent No. 6946257
; GENERAL INFORMATION:
; APPLICANT: Scott Mathew P.
; APPLICANT: Goodrich, Lisa V.
; APPLICANT: Johnson, Ronald L.
; APPLICANT: Epstein, Ervin Jr.
; TITLE OF INVENTION: PATCHED GENES AND USES RELATED THERETO
; FILE REFERENCE: CIBT-P04-203
; CURRENT APPLICATION NUMBER: US/08/916,140
; CURRENT FILING DATE: 1997-08-21
; PRIOR APPLICATION NUMBER: US 08/656,055
; PRIOR FILING DATE: 1996-05-31
; PRIOR APPLICATION NUMBER: US 08/540,406
; PRIOR FILING DATE: 1995-10-06
; PRIOR APPLICATION NUMBER: US 08/319,745
; PRIOR FILING DATE: 1994-10-07
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1447
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Primer
US-08-916-140-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy 225 PQAVGSGIQIPLNEGVARCNESQGDVATCSCDCAACPAIARPOALDSTFYLGQMPGS 284
Db 24 PGRPAGGRRRTTGLRAAAPDRD-----YLHPSYCDAAFALEQI--- 65
Qy 285 LVLIILCSVPAVVTILLVGRFVAPARDKSRMVDK-----KGTSLSDKLSFSTHLL 337
Db 66 -----SKGRATGRKAPLWLRKAFQRLLFKLCGYIQKNC 98
Qy 338 GQFFQGWGTWVASWPLTILVSVIPVALAAGLVFTELTTPDVELWSAPNSQARSEKAFH 397
Db 99 GK-----LVGLLIFGAPVGLKKAANLETNVELWVEVGGVRSELYNT 143
Qy 398 DQHFQFFRTN-QVILTAPNRSSVRYDLSLLGPKNFSGILDGLLELLE--LQERLRLH 454
Db 144 RQKIGEEAMFNQPMIQTPKEG-----ANVLTEALLQHLDSALQASRVHV 190
Qy 455 QWSPQAGRNISLQDICVAP-----LNPDTNLSYDCCINSLLQYFQNNRTLLLLTA 505
Db 191 YMYN----RQWKLHLCKYSGELITETGYMDQIIEYLPCLIIITPLDCFWECAKLQSGTA 246
Qy 506 NOTLMGO-----TSQVDWDKHFLY-----CAN----- 527
Db 247 --YLLGKPLRWTPDPLEFLEELKKNYQVDSWEMLNKAEGVGYMDRFLNAPDPDC 304
Qy 528 -----APL-----TFKQGT-----ALALSCMA 544
Db 305 PATAPKNKSTKPLDMALVINGGCHGLSRKYMHWQBELIVGGTVKNSTGKLSAHALQTMF 364
Qy 545 DYGAVPFFFLAIGGKYDKYSEAEALINTFSLNNYPAGDPRLAQAKLWEEAFLEEMRAQ 604
Db 365 QLMTPKQMY----EHFKGYEY-----VSHINWNE-----DKAAALEAWQRTYVEVHQSV 411
```

```
Qy 605 RRMAGMQVPTTAERSLEDEINRTTAEDLPFATSYIVIFLYISLALGSSYSSWRVMWDS 664
Db 412 AQNSTQKVLSEFT--TTLDDILKGSFDSVIRVASGYLLMLAYACLTM---LRWD--CSKS 465
Qy 665 KATLGLGGVAVVLGAVMAAMGFFSYLIGRSLLVTLQVVPFLVLSVGADNFIIFVLEYQRL 724
Db 466 QGAVGLAGVLLVALSVAAAGLGLSLIGISFNAATTQVLPFLALUGVGDDVFLLAHAFSE 525
Qy 725 PRRPGEPRVHIGRALGRVAPSMLLCSLSBAICFFFLGALTMPFAVTRTALTSLGVLILDF 784
Db 526 GQNKRIPIFEDRTGECLKRTCASVALTSISNVTAFFMAALIPALRAFSLQAQAVVVVNF 585
Qy 785 LLQMSAFVALLSLDSKROEASRLDVCCC-----VKQEL-----PPG 822
Db 586 AMVLLIPAILSMDLVRRDRRLDIFCCFTSPCVSRVIOVEPOAYTTHDNTRYSPPPY 645
Qy 823 QGEGL----- 827
Db 646 SSHSFAHETQITMQSTVOLTEYDPHTHYVYVYTAEPREISVQPVTVTQDTLSCQSPST 705
Qy 828 -----LLGFFQKAYAPFLLHMITRQVGVLLFLALFGVSL 861
Db 706 SSTRDLLSQFSDSLHCLPEPCTKWTLSPEAKHYAPFLKPKAKVWVIFLFLGLLQVSL 765
Qy 862 YSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGPVYFVTTGLGYNFSSEAGMNAICS 921
Db 766 YGTTVRVDGLDLDIVPRETREYDFIAAQKYSF-----YNN----- 803
Qy 922 SAGCNFSFTQKIQYAT-----EFPEQSYLAIPASS-----WVDDFIDWL----- 961
Db 804 -----YIVTQKADYPNIQHLLYDLHRSFSNVKYVMLEENKQLPKQWLHYFRDMLQGLQD 857
Qy 962 -----TFSS-----CCRLYISGNKDKFCPSTVNSLNCNKCMSITMGS 1000
Db 858 AFDSDWETGKIMPNNYKNGSDGVLAYKLLVQTGRDK--PIDISQLT--KORLVDADGI 913
Qy 1001 VRPSVEOPHKYL--PWFLND-----RPNKCPKGGGLAAV--STSVNLTSDGQ 1043
Db 914 INFSA--FYIYLTAWVSNDPVAYAASQANTRPHRPEWHDK---ADYMPETRLRIPAE 968
Qy 1044 VLASRFMAYHKPLKNSQDYTEALRAARELAANITA-DLRKVPGTDPAFEPFYTITNVFY 1102
Db 969 IEVAQPPFYLNGRLDTSDFVEAIEKVRTICSNVTSLSGLSSYPNG-----YFP---LFW 1018
Qy 1103 EQLYTLPEGLFMLSCLVETPAVSCLLGLDRLSGLNLLSIVMLIVDTVGFMAIWDIS 1162
Db 1019 EQYIGLRHLLFISVVVLAFLVCAVFLNPNWTAGII-VNVLAMTVELFGMMGLIGIK 1077
Qy 1163 YNAVSLINLYSAGVMSVEFVSHITRSP--AISTKPTWLERAKEATISMSGSAVAGVAMTN 1220
Db 1078 LSAVPVILLASVIGVEFTVHVALAFLTAIGDKN---REAVLALEHMFAPVLDG-AVST 1133
Qy 1221 LPGAIVGLAKAQLQIIFPRLNLLITLLGLHGLVFLPVLSYVG--PDVNPALALEQ- 1277
Db 1134 LLGVMLAGSEDFIVRVFFAVLAILTILGLVGLVLLPVLSFGPYPEVSPANGLNRL 1193
Qy 1278 --KRAEEAAVAAVMVASCNHPSPSVSTADNIYVNHSEFGSIKG 1317
Db 1194 PTPSPPEPPSVVRFAAMPQGH--SGSDSDSEYSSQTIVSG 1233
```

Search completed: April 7, 2006, 19:23:05  
Job time : 63 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:32:55 ; Search time 188 Seconds  
(without alignments)  
2960.365 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 MAEAGLRGWLWALLRLRLAQ.....GSIKAGATSNFLPNNGRQF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*

3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*

4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	4	US-10-621-758A-4
2	6909	100.0	1332	4	US-10-663-208A-4
3	6909	100.0	1332	4	US-10-646-301A-4
4	6909	100.0	1332	4	US-10-736-769-4
5	6896	99.8	1332	4	US-10-239-316-9
6	6872.5	99.5	1359	4	US-10-621-758A-44
7	6872.5	99.5	1359	4	US-10-663-208A-44
8	6872.5	99.5	1359	4	US-10-646-301A-44
9	6872.5	99.5	1359	4	US-10-736-769-44
10	6536	94.6	1344	5	US-10-450-763-53052
11	5421.5	78.5	1331	4	US-10-621-758A-2
12	5421.5	78.5	1331	4	US-10-663-208A-2
13	5421.5	78.5	1331	4	US-10-646-301A-2
14	5421.5	78.5	1331	4	US-10-736-769-2
15	5407	78.3	1333	4	US-10-621-758A-12
16	5407	78.3	1333	4	US-10-663-208A-12
17	5407	78.3	1333	4	US-10-646-301A-12
18	5407	78.3	1333	4	US-10-736-769-12
19	4466	64.6	982	5	US-10-450-763-53050
20	2402.5	34.8	1278	4	US-10-208-731-2
21	2402.5	34.8	1278	5	US-10-741-601-530
22	2402.5	34.8	1278	5	US-10-741-600-1542
23	2402.5	34.8	1278	5	US-10-756-149-4924
24	2385	34.5	1319	4	US-10-208-731-4
25	2262	32.7	1287	6	US-11-097-143-12003
26	1864.5	27.0	1223	6	US-11-097-143-2679
27	1329	19.2	1170	4	US-10-208-731-6

28	1078	15.6	1296	4	US-10-208-731-9	Sequence 9, Appli
29	942	13.6	194	4	US-10-239-316-8	Sequence 8, Appli
30	900.5	13.0	492	4	US-10-424-599-211862	Sequence 211862,
31	859.5	12.4	541	4	US-10-424-599-189288	Sequence 189288,
32	651.5	9.4	1447	2	US-08-954-701A-19	Sequence 19, Appli
33	651.5	9.4	1447	3	US-09-898-533-5	Sequence 5, Appli
34	651.5	9.4	1447	3	US-09-754-032-19	Sequence 19, Appli
35	651.5	9.4	1447	4	US-10-421-446-19	Sequence 19, Appli
36	651.5	9.4	1447	4	US-10-791-844-6	Sequence 6, Appli
37	646.5	9.4	1434	2	US-08-954-701A-10	Sequence 10, Appli
38	646.5	9.4	1434	3	US-09-754-032-10	Sequence 10, Appli
39	646.5	9.4	1434	4	US-10-421-446-10	Sequence 10, Appli
40	628	9.1	1296	4	US-10-302-279-60	Sequence 60, Appli
41	598.5	8.7	933	4	US-10-415-934-3	Sequence 3, Appli
42	597	8.6	821	5	US-10-890-776A-4805	Sequence 4805, Ap
43	588	8.5	954	4	US-10-060-756A-3	Sequence 3, Appli
44	588	8.5	954	5	US-10-890-776A-3	Sequence 3, Appli
45	584	8.5	1203	3	US-09-990-046-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-621-758A-4

; Sequence 4, Application US/10621758A

; Publication No. US20040093629A1

; GENERAL INFORMATION:

; APPLICANT: Altmann, Scott W

; APPLICANT: Wang, Luquan

; APPLICANT: Graziano, Michael

; APPLICANT: Murgolo, Nick

; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF

; FILE REFERENCE: JBO1603-K-US

; CURRENT APPLICATION NUMBER: US/10/621,758A

; CURRENT FILING DATE: 2003-07-17

; PRIOR APPLICATION NUMBER: 60/397,442

; PRIOR FILING DATE: 2002-07-19

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 1332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-621-758A-4

Query Match	100.0%;	Score 6909;	DB 4;	Length 1332;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1332;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MAEAGLRGWLWALLRLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGLMTLSNVSCLSN	60	
Db	1	MAEAGLRGWLWALLRLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGLMTLSNVSCLSN	60	
Qy	61	TPARKITGDHLLILQKICPRLVTGPTQACCSAKQLVSLASLSITKALLTRCPACSDNF	120	
Db	61	TPARKITGDHLLILQKICPRLVTGPTQACCSAKQLVSLASLSITKALLTRCPACSDNF	120	
Qy	121	VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFPAEQVSDSCSRVRVPA	180	
Db	121	VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFPAEQVSDSCSRVRVPA	180	
Qy	181	ATLAVGTGCVGVGSALCNAQRWLNFGDTGNGLAPLIDITFHLLEPGQAVGSGIQPLNEGV	240	
Db	181	ATLAVGTGCVGVGSALCNAQRWLNFGDTGNGLAPLIDITFHLLEPGQAVGSGIQPLNEGV	240	
Qy	241	ARCNSQGGDDVATCSQCDCAASCPAARQALDSTFYLGQMPGSLVLIILCSFAVVTI	300	
Db	241	ARCNSQGGDDVATCSQCDCAASCPAARQALDSTFYLGQMPGSLVLIILCSFAVVTI	300	
Qy	301	LLVGRVPAPARDKSNVDPKGTSLSDKLSFSTHLLQFPQCGWGTWASWPLTILVLSV	360	
Db	301	LLVGRVPAPARDKSNVDPKGTSLSDKLSFSTHLLQFPQCGWGTWASWPLTILVLSV	360	

```
QY 361 IPVVALAAGLVFTELTTPDVELWSAPNSOARSEKAFHDOHFGPPFRTNQVILTAPNRSSY 420
DB 361 IPVVALAAGLVFTELTTPDVELWSAPNSOARSEKAFHDOHFGPPFRTNQVILTAPNRSSY 420
QY 421 RYDSLGLGKPNFSGILDLDDLLLELELOERLRHLQVWSPQAQRNLSLODICYAPLNPONT 480
DB 421 RYDSLGLGKPNFSGILDLDDLLLELELOERLRHLQVWSPQAQRNLSLODICYAPLNPONT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKDHFLYCANAPLTFKDGATLAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKDHFLYCANAPLTFKDGATLAL 540
QY 541 SCMDYGAQVPPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPLAOKLWEAEFLEEM 600
DB 541 SCMDYGAQVPPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPLAOKLWEAEFLEEM 600
QY 601 RAFQRMAGMFQVFTFAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSMSRV 660
DB 601 RAFQRMAGMFQVFTFAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSMSRV 660
QY 661 MVDSKATLGLGGVAVVLGAVMAAMGFFSYLGIIRSLVILQVVPFLVLSVGADNIFIFVLE 720
DB 661 MVDSKATLGLGGVAVVLGAVMAAMGFFSYLGIIRSLVILQVVPFLVLSVGADNIFIFVLE 720
QY 721 YQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSBAICFFLIGALTPMPAVRTFALTSGLAV 780
DB 721 YQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSBAICFFLIGALTPMPAVRTFALTSGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSKQESRLDVCCVKPQELPPPGQEGEGLLGFQKAYAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSKQESRLDVCCVKPQELPPPGQEGEGLLGFQKAYAPFL 840
QY 841 LHWITRGVLLLFALFGVSLYSWCHISVGHDOELALPKDSYILDYDFLNRYPEVGAPV 900
DB 841 LHWITRGVLLLFALFGVSLYSWCHISVGHDOELALPKDSYILDYDFLNRYPEVGAPV 900
QY 901 YFVTTLGYNFSSEAGMNAICSSAGCNNEFSFTQIKYATEFPEQSYLAIPASSWDDFDW 960
DB 901 YFVTTLGYNFSSEAGMNAICSSAGCNNEFSFTQIKYATEFPEQSYLAIPASSWDDFDW 960
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSTMGSVRPSVEQPHKYLWFLNDRP 1020
DB 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSTMGSVRPSVEQPHKYLWFLNDRP 1020
QY 1021 NIKCPKGLAAYSTSVNLTSDQVLASFMAHYHPLKNSQDYTEALRAARELANITADL 1080
DB 1021 NIKCPKGLAAYSTSVNLTSDQVLASFMAHYHPLKNSQDYTEALRAARELANITADL 1080
QY 1081 RKVPCTDPAPFVFPYTTINVFYEQYLTLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140
DB 1081 RKVPCTDPAPFVFPYTTINVFYEQYLTLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140
QY 1141 NLLSIVMLVDVTFGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTWLER 1200
DB 1141 NLLSIVMLVDVTFGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTWLER 1200
QY 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQILQIIFFFELNLLITLGLHLGLVFLPV 1260
DB 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQILQIIFFFELNLLITLGLHLGLVFLPV 1260
QY 1261 ILSYVGPDPVPALAEQKRAEEAAVAVVAVASCPNHPRSVSTADNIYVNHSPFEGSIKAGA 1320
DB 1261 ILSYVGPDPVPALAEQKRAEEAAVAVVAVASCPNHPRSVSTADNIYVNHSPFEGSIKAGA 1320
QY 1321 ISNFLPNNRGQF 1332
DB 1321 ISNFLPNNRGQF 1332
```

RESULT 2  
US-10-663-208A-4  
; Sequence 4, Application US/10663208A

```
; Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC111 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J081603K2 US
; CURRENT APPLICATION NUMBER: US/10/663,208A
; PRIOR FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-663-208A-4
```

```
Query Match 100.0%; Score 6909; DB 4; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MABAGLGLWLLALLRLAQSEPYTHQPGYCAFYDECGKNPELSGLMTLSNVSCLSN 60
DB 1 MABAGLGLWLLALLRLAQSEPYTHQPGYCAFYDECGKNPELSGLMTLSNVSCLSN 60
QY 61 TPARKITGDHLIILLOKICPRLYTPGNTQACCSAKOLVSLASISITKALLTRCPACSDNF 120
DB 61 TPARKITGDHLIILLOKICPRLYTPGNTQACCSAKOLVSLASISITKALLTRCPACSDNF 120
QY 121 VNLHCHNTCPNOSLFINTVRVAQLGAGOLPAVVAYEAFYQHSFAEQSDYSCSRVRPAA 180
DB 121 VNLHCHNTCPNOSLFINTVRVAQLGAGOLPAVVAYEAFYQHSFAEQSDYSCSRVRPAA 180
QY 181 ATLAVGTMCVYGSALCNAQRWLNFGDGTNGNLAPLDITPHLLPEGOAVGSGIQPLNEGV 240
DB 181 ATLAVGTMCVYGSALCNAQRWLNFGDGTNGNLAPLDITPHLLPEGOAVGSGIQPLNEGV 240
QY 241 ARCNESQGDVATCSCQDCAASCAPATARPQALDSTFVLGMPGSLVLIILCSVEAVVTI 300
DB 241 ARCNESQGDVATCSCQDCAASCAPATARPQALDSTFVLGMPGSLVLIILCSVEAVVTI 300
QY 301 LLVGFVAPARDKSKMVDPKGTSLSDKLSFSFTHLLGQFFQGWGTWVASWPLTILVLSV 360
DB 301 LLVGFVAPARDKSKMVDPKGTSLSDKLSFSFTHLLGQFFQGWGTWVASWPLTILVLSV 360
QY 361 IPVVALAAGLVFTELTTPDVELWSAPNSOARSEKAFHDOHFGPPFRTNQVILTAPNRSSY 420
DB 361 IPVVALAAGLVFTELTTPDVELWSAPNSOARSEKAFHDOHFGPPFRTNQVILTAPNRSSY 420
QY 421 RYDSLGLGKPNFSGILDLDDLLLELELOERLRHLQVWSPQAQRNLSLODICYAPLNPONT 480
DB 421 RYDSLGLGKPNFSGILDLDDLLLELELOERLRHLQVWSPQAQRNLSLODICYAPLNPONT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKDHFLYCANAPLTFKDGATLAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVMDKDHFLYCANAPLTFKDGATLAL 540
QY 541 SCMDYGAQVPPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPLAOKLWEAEFLEEM 600
DB 541 SCMDYGAQVPPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPLAOKLWEAEFLEEM 600
QY 601 RAFQRMAGMFQVFTFAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSMSRV 660
DB 601 RAFQRMAGMFQVFTFAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSMSRV 660
QY 661 MVDSKATLGLGGVAVVLGAVMAAMGFFSYLGIIRSLVILQVVPFLVLSVGADNIFIFVLE 720
```

Db 661 MVDKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLQVVPFLVSLVGADNIFIFVLE 720  
QY 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAI CFFLGALTMPAVRTTALTSGLAV 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAI CFFLGALTMPAVRTTALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKQASRLDVCCCVKQPQLPPPGQEGLLLGFFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQASRLDVCCCVKQPQLPPPGQEGLLLGFFQKAYAPFL 840  
QY 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
QY 901 YFVTTLGYNFSSEAGNNAICSSAGCANNFSTQKIQYATFPPEOSYLAIPASSWVDDFIDW 960  
Db 901 YFVTTLGYNFSSEAGNNAICSSAGCANNFSTQKIQYATFPPEOSYLAIPASSWVDDFIDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQFHKYL PWF LND RP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQFHKYL PWF LND RP 1020  
QY 1021 NIKCPKGLAAYSTSVNLSDGQVLASRFMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLSDGQVLASRFMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
QY 1081 RKVPGTDPAPFVFPYITVNFVEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAPFVFPYITVNFVEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140  
QY 1141 NLLSIVMLVDTVGFMALWDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSIVMLVDTVGFMALWDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMSAVFAGVAMTNLGIILVGLAKAQLIQIIPFRLNLLITLGLLHGLVFLPV 1260  
Db 1201 AKEATISMSAVFAGVAMTNLGIILVGLAKAQLIQIIPFRLNLLITLGLLHGLVFLPV 1260  
QY 1261 ILSYVGPDVNPALAEQKRAEBAVAAMVAVASCPNHPSPRVSTADNIVNHSFPGSIKGAGA 1320  
Db 1261 ILSYVGPDVNPALAEQKRAEBAVAAMVAVASCPNHPSPRVSTADNIVNHSFPGSIKGAGA 1320  
QY 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

RESULT 3

US-10-646-301A-4  
; Sequence 4, Application US/10646301A  
; Publication No. US20040137467A1  
; GENERAL INFORMATION:  
; APPLICANT: Altman, Scott W  
; APPLICANT: Wang, Luquan  
; APPLICANT: Graziano, Michael  
; APPLICANT: Murgolo, Nick  
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF  
; FILE REFERENCE: JB01603-K1-US  
; CURRENT APPLICATION NUMBER: US/10/646, 301A  
; CURRENT FILING DATE: 2003-08-22  
; PRIOR APPLICATION NUMBER: 60/397,442  
; PRIOR FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 10/621,758  
; PRIOR FILING DATE: 2003-07-17  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-646-301A-4

Query Match 100.0%; Score 6909; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEAGLRGWLWALLRLLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCSLSN 60  
Db 1 MAEAGLRGWLWALLRLLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCSLSN 60  
QY 61 TPARKITGDHLLILLOKICPRLTYGPNTOACCSAKQLVSLSEASITIKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLTYGPNTOACCSAKQLVSLSEASITIKALLTRCPACSDNF 120  
QY 121 VNLHCHTCSPNQSLFINVTRVAQAGOLPAVAYEAFYQHSFABQSYDSCSRVRPAA 180  
Db 121 VNLHCHTCSPNQSLFINVTRVAQAGOLPAVAYEAFYQHSFABQSYDSCSRVRPAA 180  
QY 181 ATLAGVTGCVYGSALCNAQRWLNFGDTGNGLAPLDITFHLLERQOAVSGIOPRNEG 240  
Db 181 ATLAGVTGCVYGSALCNAQRWLNFGDTGNGLAPLDITFHLLERQOAVSGIOPRNEG 240  
QY 241 ARCNESSQDDVATCSCQDCAASCPAIPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300  
Db 241 ARCNESSQDDVATCSCQDCAASCPAIPQALDSTFYLGQMPGSLVLIILICSVFAVVTI 300  
QY 301 LLVGRFVAPARDKSKMDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVASWPLTILVLSV 360  
Db 301 LLVGRFVAPARDKSKMDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVASWPLTILVLSV 360  
QY 361 IPVVALAAGLVTELTTPDVELWSAPNSQARSEKAFHDQHFPGFRTNQVILTAPNRSSY 420  
Db 361 IPVVALAAGLVTELTTPDVELWSAPNSQARSEKAFHDQHFPGFRTNQVILTAPNRSSY 420  
QY 421 RYDSSLGPKPNSGILDLDLLELEQLERLHLQVWSPQAORNTSLQDICIAPLNPDNT 480  
Db 421 RYDSSLGPKPNSGILDLDLLELEQLERLHLQVWSPQAORNTSLQDICIAPLNPDNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANTQTLMGQTSQVDWKDHFLYCANAPLTFKDGTALAL 540  
Db 481 SLYDCCINSLLQYFQNNRTLLLTANTQTLMGQTSQVDWKDHFLYCANAPLTFKDGTALAL 540  
QY 541 SCMDYGPVFPFLAIGYKGDYSEAEALIMTFSLNYPAGDPRLAOKLWEEAFLEEM 600  
Db 541 SCMDYGPVFPFLAIGYKGDYSEAEALIMTFSLNYPAGDPRLAOKLWEEAFLEEM 600  
QY 601 RAFORMAGMFOVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660  
Db 601 RAFORMAGMFOVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660  
QY 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLQVVPFLVSLVGADNIFIFVLE 720  
Db 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLQVVPFLVSLVGADNIFIFVLE 720  
QY 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAI CFFLGALTMPAVRTTALTSGLAV 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMLLCSLSEAI CFFLGALTMPAVRTTALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKQASRLDVCCCVKQPQLPPPGQEGLLLGFFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQASRLDVCCCVKQPQLPPPGQEGLLLGFFQKAYAPFL 840  
QY 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
QY 901 YFVTTLGYNFSSEAGNNAICSSAGCANNFSTQKIQYATFPPEOSYLAIPASSWVDDFIDW 960  
Db 901 YFVTTLGYNFSSEAGNNAICSSAGCANNFSTQKIQYATFPPEOSYLAIPASSWVDDFIDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQFHKYL PWF LND RP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMSITMGSVRPSVEQFHKYL PWF LND RP 1020  
QY 1021 NIKCPKGLAAYSTSVNLSDGQVLASRFMAYHKPLKNSQDYTEALRAARELAANITADL 1080

Db 1021 NIKCPKGLAAVSTSVNLTSQGVLASRFMAVHKPLKNSQDYTEALRAARELANITADL 1080  
QY 1081 RKVPGTDPAFVFPYTTINVFYEQYLITLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFVFPYTTINVFYEQYLITLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140  
QY 1141 NLLSVMLVLDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSVMLVLDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
QY 1261 ILSYVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSEFGSIKGAGA 1320  
Db 1261 ILSYVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSEFGSIKGAGA 1320  
QY 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

## RESULT 4

US-10-736-769-4

; Sequence 4, Application US/10736769

; Publication No. US20040161838A1

; GENERAL INFORMATION:

; APPLICANT: Altmann, Scott W

; APPLICANT: Wang, Luquan

; APPLICANT: Graziano, Michael

; APPLICANT: Murgolo, Nick

; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF

; FILE REFERENCE: JBO1603-K3-US

; CURRENT APPLICATION NUMBER: US/10/736,769

; CURRENT FILING DATE: 2003-12-16

; PRIOR FILING DATE: 2003-07-19

; PRIOR FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 10/621,758

; PRIOR FILING DATE: 2003-07-17

; PRIOR APPLICATION NUMBER: 10/646,301

; PRIOR FILING DATE: 2003-08-22

; PRIOR APPLICATION NUMBER: 10/663,208

; PRIOR FILING DATE: 2003-09-16

; NUMBER OF SEQ ID NOS: 51

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 1332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-736-769-4

Query Match 100.0%; Score 6909; DB 4; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLRCGLLWALLRLRAQSEPYTTIHQPVCAPYDECCKNPGLSGSLMTLSNVSCLSN 60  
Db 1 MAEAGLRCGLLWALLRLRAQSEPYTTIHQPVCAPYDECCKNPGLSGSLMTLSNVSCLSN 60  
QY 61 TPARKITGDHLILLOKICPRLYTGPTNTQACCSAKQLVSEASLSITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLILLOKICPRLYTGPTNTQACCSAKQLVSEASLSITKALLTRCPACSDNF 120  
QY 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFQHSFABQSYDCSRRVPAA 180  
Db 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFQHSFABQSYDCSRRVPAA 180  
QY 181 ATLAAGTMCVGYGAGLCAQWLNFGQDGTGNGLAPLDTITFHLLPFGQAVGSGIOPLEGV 240  
Db 181 ATLAAGTMCVGYGAGLCAQWLNFGQDGTGNGLAPLDTITFHLLPFGQAVGSGIOPLEGV 240

QY 241 ARCNESGDDVATCSODCAASCPATARPQALDSTFYLGOMPSGLVLIILILCSVFAVVTI 300  
Db 241 ARCNESGDDVATCSODCAASCPATARPQALDSTFYLGOMPSGLVLIILILCSVFAVVTI 300  
QY 301 LLVGFVRVAPARDSKMWDPKGTSLSKLSFSFTHLLGQFFQFGMGTVWASWPLTILVLSV 360  
Db 301 LLVGFVRVAPARDSKMWDPKGTSLSKLSFSFTHLLGQFFQFGMGTVWASWPLTILVLSV 360  
QY 361 IPVVALAAGLVFTTELTTDPVELWSAPNSQARSKAHQDHQFGFPFRFNQVILTAPNRSY 420  
Db 361 IPVVALAAGLVFTTELTTDPVELWSAPNSQARSKAHQDHQFGFPFRFNQVILTAPNRSY 420  
QY 421 RYDSSLILGPKNFSGILDLLELLELLELLELLELLELLELLELLELLELLELLELLELLEL 480  
Db 421 RYDSSLILGPKNFSGILDLLELLELLELLELLELLELLELLELLELLELLELLELLEL 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKDGTLAL 540  
Db 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKDGTLAL 540  
QY 541 SCHADYGAPVFPFLATCGYKGOYSEBALIMTFSLNNYPAGDPRLAQAQKLWBEAFLEEM 600  
Db 541 SCHADYGAPVFPFLATCGYKGOYSEBALIMTFSLNNYPAGDPRLAQAQKLWBEAFLEEM 600  
QY 601 RAFORMAGMFQVTTAERSLEDEINRTTAEDLPIFATSIVIFLYXISLALGSVSSMSRV 660  
Db 601 RAFORMAGMFQVTTAERSLEDEINRTTAEDLPIFATSIVIFLYXISLALGSVSSMSRV 660  
QY 661 MVDSKATLGLGGVAVVLGAVMAAMGFFSYIGIRSSLVILQVVPFLVLSGADNIFIFVLE 720  
Db 661 MVDSKATLGLGGVAVVLGAVMAAMGFFSYIGIRSSLVILQVVPFLVLSGADNIFIFVLE 720  
QY 721 YQRLPRRPGEPREVHIGRALGRVAPSMILCSLSEATCFPLGALTTPMPAVRTFALTSGLA 780  
Db 721 YQRLPRRPGEPREVHIGRALGRVAPSMILCSLSEATCFPLGALTTPMPAVRTFALTSGLA 780  
QY 781 ILDFLLQMSAFVALLSLDSKRQASRLDVCCCKVPOELPPPGQEGLLGFFOKAYAPPL 840  
Db 781 ILDFLLQMSAFVALLSLDSKRQASRLDVCCCKVPOELPPPGQEGLLGFFOKAYAPPL 840  
QY 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
QY 901 YFVTTILGYNFSSEAGNNAICSSAGCNNEFPTQIOYATEPEQSYLAIPASSWVDIDW 960  
Db 901 YFVTTILGYNFSSEAGNNAICSSAGCNNEFPTQIOYATEPEQSYLAIPASSWVDIDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSITMGSVRSPSVQFHKYLPWFNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSITMGSVRSPSVQFHKYLPWFNDRP 1020  
QY 1021 NIKCPKGLAAVSTSVNLTSQGVLASRFMAVHKPLKNSQDYTEALRAARELANITADL 1080  
Db 1021 NIKCPKGLAAVSTSVNLTSQGVLASRFMAVHKPLKNSQDYTEALRAARELANITADL 1080  
QY 1081 RKVPGTDPAFVFPYTTINVFYEQYLITLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFVFPYTTINVFYEQYLITLPEGLFMLSICLVPTFAVSCLLGLDLRSGLL 1140  
QY 1141 NLLSVMLVLDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSVMLVLDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
QY 1261 ILSYVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSEFGSIKGAGA 1320  
Db 1261 ILSYVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSEFGSIKGAGA 1320  
QY 1321 ISNFLPNNGRQF 1332





```
Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy 1 MAEAGLRGWLWALLRLAQSSEPYTHHOPGYCAFYDECCKNPGLSGMLTSLNVCSLSN 60
Db 1 MAEAGLRGWLWALLRLAQSSEPYTHHOPGYCAFYDECCKNPGLSGMLTSLNVCSLSN 60
Qy 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSEASLSITKALLTRCPACSDNF 120
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSEASLSITKALLTRCPACSDNF 120
Qy 121 VNLHCHNTCSNQSFLINVTVAQLGAGQLPAVVAEAFYQHSFAEQSYDCSRVRPAA 180
Db 121 VNLHCHNTCSNQSFLINVTVAQLGAGQLPAVVAEAFYQHSFAEQSYDCSRVRPAA 180
Qy 181 ATAVGTWCGVYSGALCNAORWLNFGDGTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240
Db 181 ATAVGTWCGVYSGALCNAORWLNFGDGTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240
Qy 241 ARCNEQGGDDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
Db 241 ARCNEQGGDDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
Qy 301 LLVGRVAPARDKSNWDPKGTSLSDKLSFTHTLLGQFQCGTGWVASHPLILVLSV 360
Db 301 LLVGRVAPARDKSNWDPKGTSLSDKLSFTHTLLGQFQCGTGWVASHPLILVLSV 360
Qy 361 IPVVALAAGLVFTELTTDPVELWAPNSQARSEKAFHDHGFPPFRFTNOVILTAPNRSY 420
Db 361 IPVVALAAGLVFTELTTDPVELWAPNSQARSEKAFHDHGFPPFRFTNOVILTAPNRSY 420
Qy 421 RYDSLLGPKNFSGLDLDLLELLELQERLHQLVMSPEAQRNISLQDICYAPLNPNT 480
Db 421 RYDSLLGPKNFSGLDLDLLELLELQERLHQLVMSPEAQRNISLQDICYAPLNPNT 480
Qy 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFYCANAPLTFKDGTLAL 540
Db 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFYCANAPLTFKDGTLAL 540
Qy 541 SCMDYAGAPVPFFLAIGYKGDYSEAEALIMTSLNYPAGDPRLAQAKLWEAEFLM 600
Db 541 SCMDYAGAPVPFFLAIGYKGDYSEAEALIMTSLNYPAGDPRLAQAKLWEAEFLM 600
Qy 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSWSRV 660
Db 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSWSRV 660
Qy 661 MVDKATIGLGVAVVLGAVNAAMGFYSYLGIRSLVTLQVVPFLVLSVGDNIFIFVLE 720
Db 661 MVDKATIGLGVAVVLGAVNAAMGFYSYLGIRSLVTLQVVPFLVLSVGDNIFIFVLE 720
Qy 721 YQRLPRRPGEPREHVHIGALGRVAPSMLLCSLSEALCFFLGALTMPAVRTFALTSLGLAV 780
Db 721 YQRLPRRPGEPREHVHIGALGRVAPSMLLCSLSEALCFFLGALTMPAVRTFALTSLGLAV 780
Qy 781 ILDFLLQMSAFVALLSLDSKQEBASRLDVCCVCPQELPPPGQEGLLLGFFQKAYAPFL 840
Db 781 ILDFLLQMSAFVALLSLDSKQEBASRLDVCCVCPQELPPPGQEGLLLGFFQKAYAPFL 840
Qy 841 LHWITRGVVLFLFALFGVLSYSMCHISVGLDQELALPKDSYLLDYFLFLNRYEFGAPV 900
Db 841 LHWITRGVVLFLFALFGVLSYSMCHISVGLDQELALPKDSYLLDYFLFLNRYEFGAPV 900
Qy 901 YFVTTLGYNFSSAGMNAICSSAGCNFSPFKIQIYATEFPEQSYLAIPASSWDDFDW 960
Db 901 YFVTTLGYNFSSAGMNAICSSAGCNFSPFKIQIYATEFPEQSYLAIPASSWDDFDW 960
Qy 961 LTPSSCCRLYISGNKDKFCSTVNSLNCNCSITWGSVRPSVEQPHKLPWFINDRP 1020
Db 961 LTPSSCCRLYISGNKDKFCSTVNSLNCNCSITWGSVRPSVEQPHKLPWFINDRP 1020

Qy 1021 NIKCPKGLAAYSTSVNLTSDGQVL-----ASRFMAYH 1053
Db 1021 NIKCPKGLAAYSTSVNLTSDGQVLDTVAITLSPRLEYSGTISAHCNLYLSDSASRFMAYH 1080
Qy 1054 KPLKNSQDYTEALARAARELAANITADLRKVPGTDPAFEPFPTTNNVFYEQYLTILPEGL 1113
Db 1054 KPLKNSQDYTEALARAARELAANITADLRKVPGTDPAFEPFPTTNNVFYEQYLTILPEGL 1113
Qy 1081 KPLKNSQDYTEALARAARELAANITADLRKVPGTDPAFEPFPTTNNVFYEQYLTILPEGL 1140
Db 1081 KPLKNSQDYTEALARAARELAANITADLRKVPGTDPAFEPFPTTNNVFYEQYLTILPEGL 1140
Qy 1114 FMLSCLVPTTFVAVSCILLGLDLRSLNLLSIYVILVDTVGFPMALMDISYNAVSLINLVS 1173
Db 1114 FMLSCLVPTTFVAVSCILLGLDLRSLNLLSIYVILVDTVGFPMALMDISYNAVSLINLVS 1173
Qy 1141 FMLSCLVPTTFVAVSCILLGLDLRSLNLLSIYVILVDTVGFPMALMDISYNAVSLINLVS 1200
Db 1141 FMLSCLVPTTFVAVSCILLGLDLRSLNLLSIYVILVDTVGFPMALMDISYNAVSLINLVS 1200
Qy 1174 AVGMSVEFVSHITRFAISTKPTWLRBRKEATISMGSAVPAVGMNTLPGILVLGLAKAQ 1233
Db 1174 AVGMSVEFVSHITRFAISTKPTWLRBRKEATISMGSAVPAVGMNTLPGILVLGLAKAQ 1233
Qy 1201 AVGMSVEFVSHITRFAISTKPTWLRBRKEATISMGSAVPAVGMNTLPGILVLGLAKAQ 1260
Db 1201 AVGMSVEFVSHITRFAISTKPTWLRBRKEATISMGSAVPAVGMNTLPGILVLGLAKAQ 1260
Qy 1234 LIQIFFRNLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALALQKRAEEAAVAAVMVASCP 1293
Db 1234 LIQIFFRNLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALALQKRAEEAAVAAVMVASCP 1293
Qy 1261 LIQIFFRNLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALALQKRAEEAAVAAVMVASCP 1320
Db 1261 LIQIFFRNLNLLITLGLLHGLVFLPVILSYVGPVDPNPAALALQKRAEEAAVAAVMVASCP 1320
Qy 1294 NHPSRVSTADNIYVNHSEFSGIKGAGAINFLPNNGRQF 1332
Db 1321 NHPSRVSTADNIYVNHSEFSGIKGAGAINFLPNNGRQF 1359

RESULT 7
US-10-663-208A-44
; Sequence 44, Application US/10663208A
; Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603K2 US
; CURRENT APPLICATION NUMBER: US/10/663, 208A
; CURRENT FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-663-208A-44

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy 1 MAEAGLRGWLWALLRLAQSSEPYTHHOPGYCAFYDECCKNPGLSGMLTSLNVCSLSN 60
Db 1 MAEAGLRGWLWALLRLAQSSEPYTHHOPGYCAFYDECCKNPGLSGMLTSLNVCSLSN 60
Qy 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSEASLSITKALLTRCPACSDNF 120
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSEASLSITKALLTRCPACSDNF 120
Qy 121 VNLHCHNTCSNQSFLINVTVAQLGAGQLPAVVAEAFYQHSFAEQSYDCSRVRPAA 180
Db 121 VNLHCHNTCSNQSFLINVTVAQLGAGQLPAVVAEAFYQHSFAEQSYDCSRVRPAA 180
Qy 181 ATAVGTWCGVYSGALCNAORWLNFGDGTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240
Db 181 ATAVGTWCGVYSGALCNAORWLNFGDGTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240
Qy 241 ARCNEQGGDDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
Db 241 ARCNEQGGDDVATCSCQDCAASCPAIAARPOALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
Qy 301 LLVGRVAPARDKSNWDPKGTSLSDKLSFTHTLLGQFQCGTGWVASHPLILVLSV 360
Db 301 LLVGRVAPARDKSNWDPKGTSLSDKLSFTHTLLGQFQCGTGWVASHPLILVLSV 360
Qy 361 IPVVALAAGLVFTELTTDPVELWAPNSQARSEKAFHDHGFPPFRFTNOVILTAPNRSY 420
Db 361 IPVVALAAGLVFTELTTDPVELWAPNSQARSEKAFHDHGFPPFRFTNOVILTAPNRSY 420
Qy 421 RYDSLLGPKNFSGLDLDLLELLELQERLHQLVMSPEAQRNISLQDICYAPLNPNT 480
Db 421 RYDSLLGPKNFSGLDLDLLELLELQERLHQLVMSPEAQRNISLQDICYAPLNPNT 480
Qy 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFYCANAPLTFKDGTLAL 540
Db 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVDWKDHFYCANAPLTFKDGTLAL 540
Qy 541 SCMDYAGAPVPFFLAIGYKGDYSEAEALIMTSLNYPAGDPRLAQAKLWEAEFLM 600
Db 541 SCMDYAGAPVPFFLAIGYKGDYSEAEALIMTSLNYPAGDPRLAQAKLWEAEFLM 600
Qy 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSWSRV 660
Db 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSYSWSRV 660
Qy 661 MVDKATIGLGVAVVLGAVNAAMGFYSYLGIRSLVTLQVVPFLVLSVGDNIFIFVLE 720
Db 661 MVDKATIGLGVAVVLGAVNAAMGFYSYLGIRSLVTLQVVPFLVLSVGDNIFIFVLE 720
Qy 721 YQRLPRRPGEPREHVHIGALGRVAPSMLLCSLSEALCFFLGALTMPAVRTFALTSLGLAV 780
Db 721 YQRLPRRPGEPREHVHIGALGRVAPSMLLCSLSEALCFFLGALTMPAVRTFALTSLGLAV 780
Qy 781 ILDFLLQMSAFVALLSLDSKQEBASRLDVCCVCPQELPPPGQEGLLLGFFQKAYAPFL 840
Db 781 ILDFLLQMSAFVALLSLDSKQEBASRLDVCCVCPQELPPPGQEGLLLGFFQKAYAPFL 840
Qy 841 LHWITRGVVLFLFALFGVLSYSMCHISVGLDQELALPKDSYLLDYFLFLNRYEFGAPV 900
Db 841 LHWITRGVVLFLFALFGVLSYSMCHISVGLDQELALPKDSYLLDYFLFLNRYEFGAPV 900
Qy 901 YFVTTLGYNFSSAGMNAICSSAGCNFSPFKIQIYATEFPEQSYLAIPASSWDDFDW 960
Db 901 YFVTTLGYNFSSAGMNAICSSAGCNFSPFKIQIYATEFPEQSYLAIPASSWDDFDW 960
Qy 961 LTPSSCCRLYISGNKDKFCSTVNSLNCNCSITWGSVRPSVEQPHKLPWFINDRP 1020
Db 961 LTPSSCCRLYISGNKDKFCSTVNSLNCNCSITWGSVRPSVEQPHKLPWFINDRP 1020
```



Db	241	ARCNESQGDVATCSQDCCAASCPAIA	RPQALDSTFYLGQMPGSLVLIILICSVFAVVTI	300
Qy	301	LLVGRVAPARDKSMVDPKKGTSLSDKLSFS	THLLGQFFQGGTGWASWPLTILVLSV	360
Db	301	LLVGRVAPARDKSMVDPKKGTSLSDKLSFS	THLLGQFFQGGTGWASWPLTILVLSV	360
Qy	361	IPVVALAAGLVFTELTDPVELWSAPNSQAR	SKAFHQHFGFFRTNQVILTPAPNRSSY	420
Db	361	IPVVALAAGLVFTELTDPVELWSAPNSQAR	SKAFHQHFGFFRTNQVILTPAPNRSSY	420
Qy	421	RYDSLILGPKNFGSILDLLELLELQERLHL	QVWSPEAQRNLSLODICYAPLNPNT	480
Db	421	RYDSLILGPKNFGSILDLLELLELQERLHL	QVWSPEAQRNLSLODICYAPLNPNT	480
Qy	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMG	QTSQVDWKDHFLYCANAPLTFKDGTTALAL	540
Db	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMG	QTSQVDWKDHFLYCANAPLTFKDGTTALAL	540
Qy	541	SCMADYGAPVPPFLAIGYKGDYSEABALIM	TFSLNNYPAGDPRPRLAQAKLWEEAFLEEM	600
Db	541	SCMADYGAPVPPFLAIGYKGDYSEABALIM	TFSLNNYPAGDPRPRLAQAKLWEEAFLEEM	600
Qy	601	RAFQRMAGMFOVTFTAERSLEDEINRTAED	LPIFATSYIVIFLYISALGSYSWSRV	660
Db	601	RAFQRMAGMFOVTFTAERSLEDEINRTAED	LPIFATSYIVIFLYISALGSYSWSRV	660
Qy	661	MVDSKATLGLGGVAVVLGVAWMAAGPFSY	LGRSSVLVQVVPFLVLSVGADNIFVLE	720
Db	661	MVDSKATLGLGGVAVVLGVAWMAAGPFSY	LGRSSVLVQVVPFLVLSVGADNIFVLE	720
Qy	721	YQRLPRRPGEPREHIGRALGRVAPSMLLCS	LSEAI CFFLGALTMPAVRTFALTSGLAV	780
Db	721	YQRLPRRPGEPREHIGRALGRVAPSMLLCS	LSEAI CFFLGALTMPAVRTFALTSGLAV	780
Qy	781	ILDFLQMSAFVALLSDSKRQASRLDVCCCK	PQELPPPGQEGILLIGFFQKAYAPFL	840
Db	781	ILDFLQMSAFVALLSDSKRQASRLDVCCCK	PQELPPPGQEGILLIGFFQKAYAPFL	840
Qy	841	LHWITRGVLLFLALFGVSLYSWCHISVGLD	QELALPKDSYLLDYFLFLNRYFEVGA	900
Db	841	LHWITRGVLLFLALFGVSLYSWCHISVGLD	QELALPKDSYLLDYFLFLNRYFEVGA	900
Qy	901	YFVTTLGYNFSSEAGNNAICSSAGC	NNFSTQKIQYATEFPEQSYLAIIPASS	960
Db	901	YFVTTLGYNFSSEAGNNAICSSAGC	NNFSTQKIQYATEFPEQSYLAIIPASS	960
Qy	961	LTPSSCCRLYISGPNKDKFCPTVNSL	NCLKNCMSITMGSVRPSVQFHKL	1020
Db	961	LTPSSCCRLYISGPNKDKFCPTVNSL	NCLKNCMSITMGSVRPSVQFHKL	1020
Qy	1021	NIKCPKGGLAAYSTSVNLTSDGOVL	-----ASRFMAVH	1053
Db	1021	NIKCPKGGLAAYSTSVNLTSDGOVL	-----ASRFMAVH	1053
Qy	1054	KPLKNSQDYTEALRAARELAANITADLR	KVPDPAFEVFPYITNVFYEQYLTILPEGL	1113
Db	1081	KPLKNSQDYTEALRAARELAANITADLR	KVPDPAFEVFPYITNVFYEQYLTILPEGL	1140
Qy	1114	FMLSCLVPTFAVSCLLIGDLASGLNLS	IVMILVDTVGFMAWMDISYNAVSLINLVS	1173
Db	1141	FMLSCLVPTFAVSCLLIGDLASGLNLS	IVMILVDTVGFMAWMDISYNAVSLINLVS	1200
Qy	1174	AVGMSVEFYSHITRSFAITKPTWLERAKE	ATISMGSAVFAVAMTNLPGLVLGLAKAQ	1233
Db	1201	AVGMSVEFYSHITRSFAITKPTWLERAKE	ATISMGSAVFAVAMTNLPGLVLGLAKAQ	1260
Qy	1234	LIQIFPFRNLNLTLLGLHLGVFLFVL	SVILSYVGPVNPALALEQKAEAEAAVWV	1293
Db	1261	LIQIFPFRNLNLTLLGLHLGVFLFVL	SVILSYVGPVNPALALEQKAEAEAAVW	1320
Qy	1294	NHPSRVSTADNIYVNHSPFGS	IKGAGAIISNFIPLNNGRQF	1332

Db	1321	NHPSRVSTADNIYVNHSPFGS	IKGAGAIISNFIPLNNGRQF	1359
RESULT 8				
US-10-646-301A-44				
; Sequence 44, Application US/10646301A				
; Publication No. US20040137467A1				
; GENERAL INFORMATION:				
; APPLICANT: Altmann, Scott W				
; APPLICANT: Wang, Luquan				
; APPLICANT: Graziano, Michael				
; APPLICANT: Murgolo, Nick				
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF				
; FILE REFERENCE: JBO1603-K1-US				
; CURRENT APPLICATION NUMBER: US/10/646,301A				
; CURRENT FILING DATE: 2003-08-22				
; PRIOR FILING DATE: 2002-07-19				
; PRIOR APPLICATION NUMBER: 60/397,442				
; PRIOR FILING DATE: 2003-07-17				
; NUMBER OF SEQ ID NOS: 50				
; SOFTWARE: Patentin version 3.1				
; SEQ ID NO 44				
; LENGTH: 1359				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-10-646-301A-44				
Query Match 99.5%; Score 6872.5; DB 4; Length 1359;				
Best Local Similarity 97.9%; Pred. No. 0;				
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;				
Qy	1	MAEAGLRGMLLWALLRLAQSEPYTTIHOPGY	CAFYDECGKPELSGSLMTLSNVCSLSN	60
Db	1	MAEAGLRGMLLWALLRLAQSEPYTTIHOPGY	CAFYDECGKPELSGSLMTLSNVCSLSN	60
Qy	61	TPARKITGDHLLILQKICPRLYTGPNTQACCS	AKQLVLSLEASISITKALLTRCPACSDNF	120
Db	61	TPARKITGDHLLILQKICPRLYTGPNTQACCS	AKQLVLSLEASISITKALLTRCPACSDNF	120
Qy	121	VNLHCHNTCSNQSLFINVTRVAQLGAGOLPA	WAVEAYEAFYQHSFABQSDSCSRVRPAA	180
Db	121	VNLHCHNTCSNQSLFINVTRVAQLGAGOLPA	WAVEAYEAFYQHSFABQSDSCSRVRPAA	180
Qy	181	ATLAVGTMGVYGSALCNAQRWLNFGDGTG	NGLAPLDTTFHLLPEQOAVGSGIQPLNEGV	240
Db	181	ATLAVGTMGVYGSALCNAQRWLNFGDGTG	NGLAPLDTTFHLLPEQOAVGSGIQPLNEGV	240
Qy	241	ARCNESQGDVATCSQDCAASC	PAIARPAQLDSTFYLGQMPGSLVLIILICSVFAVVTI	300
Db	241	ARCNESQGDVATCSQDCAASC	PAIARPAQLDSTFYLGQMPGSLVLIILICSVFAVVTI	300
Qy	301	LLVGRVAPARDKSKMVDPKKGTSLSDKLS	FSFTHLLGQFFQGGTGWASWPLTILVLSV	360
Db	301	LLVGRVAPARDKSKMVDPKKGTSLSDKLS	FSFTHLLGQFFQGGTGWASWPLTILVLSV	360
Qy	361	IPVVALAAGLVFTELTDPVELWSAPNSQAR	SEKAFHQHFGFFRTNQVILTPAPNRSSY	420
Db	361	IPVVALAAGLVFTELTDPVELWSAPNSQAR	SEKAFHQHFGFFRTNQVILTPAPNRSSY	420
Qy	421	RYDSLILGPKNFGSILDLLELLELQERLHL	QVWSPEAQRNLSLODICYAPLNPNT	480
Db	421	RYDSLILGPKNFGSILDLLELLELQERLHL	QVWSPEAQRNLSLODICYAPLNPNT	480
Qy	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMG	QTSQVDWKDHFLYCANAPLTFKDGTTALAL	540
Db	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMG	QTSQVDWKDHFLYCANAPLTFKDGTTALAL	540
Qy	541	SCMADYGAPVPPFLAIGYKGDYSEABALIM	TFSLNNYPAGDPRPRLAQAKLWEEAFLEEM	600
Db	541	SCMADYGAPVPPFLAIGYKGDYSEABALIM	TFSLNNYPAGDPRPRLAQAKLWEEAFLEEM	600
Qy	601	RAFQRMAGMFOVTFTAERSLEDEINRTAED	LPIFATSYIVIFLYISALGSYSWSRV	660

Db 601 RAQRMAGNFQVTFMAERSLEDEINRTTAEDLPFATSYIVIFLYISIALGYSWSRV 660  
Qy 661 MVDKATLGLGGVAVVLGAVMAAMGFSSYLGRSSVLIVQVPPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVVLGAVMAAMGFSSYLGRSSVLIVQVPPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSEACFFLGTALTPMPAVRTFALTSGLA 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSEACFFLGTALTPMPAVRTFALTSGLA 780  
Qy 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKVQDELALPKDSYLLDYFLFLNRYFEVGAPV 840  
Db 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKVQDELALPKDSYLLDYFLFLNRYFEVGAPV 840  
Qy 841 LHWITRGVVLILLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVVLILLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Qy 901 YFVTTILGYNFSSSEAGNNAICSSAGCANNFSTOKI QYATEFFPROSYLAIPASSWVDDFDW 960  
Db 901 YFVTTILGYNFSSSEAGNNAICSSAGCANNFSTOKI QYATEFFPROSYLAIPASSWVDDFDW 960  
Qy 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCNKMSITMGSVRPSVEQPHKYLPPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCNKMSITMGSVRPSVEQPHKYLPPWFLNDRP 1020  
Qy 1021 NIKCPKGGGLAAVSTSNLSDGVL-----ASRFMAYH 1053  
Db 1021 NIKCPKGGGLAAVSTSNLSDGQVLDTVAILSPRLEYSGTISAHCNLYLLDSASRFMAYH 1080  
Qy 1054 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYTIITNVFYEQVLTILPEGL 1113  
Db 1081 KPLKNSQDYTEALRAARELAANITADLRKVPCTDPAFEVFPYTIITNVFYEQVLTILPEGL 1140  
Qy 1114 FMLSCLVPTFAVSCLLGLDLSGLNLLSYWILVDTVGPMAWLDISYNAVSLINLVS 1173  
Db 1141 FMLSCLVPTFAVSCLLGLDLSGLNLLSYWILVDTVGPMAWLDISYNAVSLINLVS 1200  
Qy 1174 AVGMSVEFVSHITRFAISTKFTWLERAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1233  
Db 1201 AVGMSVEFVSHITRFAISTKFTWLERAKEATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1260  
Qy 1234 LIQIFFFRLNLITLLGLHLGVFLPVILSYVGPDPVPALALEQKRAEEA VAAVWVASC 1293  
Db 1261 LIQIFFFRLNLITLLGLHLGVFLPVILSYVGPDPVPALALEQKRAEEA VAAVWVASC 1320  
Qy 1294 NHPSRVSTADNIYVNHSPFSGSTKGAGAINFLPNNGROF 1332  
Db 1321 NHPSRVSTADNIYVNHSPFSGSTKGAGAINFLPNNGROF 1359

## RESULT 9

US-10-736-769-44  
; Sequence 44, Application US/10736769  
; Publication No. US20040161838A1  
; GENERAL INFORMATION:  
; APPLICANT: Altman, Scott W  
; APPLICANT: Wang, Luquan  
; APPLICANT: Graziano, Michael  
; APPLICANT: Murgolo, Nick  
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF  
; FILE REFERENCE: JB01603-K3-US  
; CURRENT APPLICATION NUMBER: US/10/736,769  
; CURRENT FILING DATE: 2003-12-16  
; PRIOR APPLICATION NUMBER: 60/397,442  
; PRIOR FILING DATE: 2002-07-19  
; PRIOR APPLICATION NUMBER: 10/621,758  
; PRIOR FILING DATE: 2003-07-17  
; PRIOR APPLICATION NUMBER: 10/646,301  
; PRIOR FILING DATE: 2003-08-22  
; PRIOR APPLICATION NUMBER: 10/663,208  
; PRIOR FILING DATE: 2003-09-16

; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 44  
; LENGTH: 1359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-736-769-44

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy 1 MAEAGLIGWLLWALLLLBLAQSEPYTHIQPGYCAFYDECCKNPDELSSGLMTLSNVSCLSN 60  
Db 1 MAEAGLIGWLLWALLLLBLAQSEPYTHIQPGYCAFYDECCKNPDELSSGLMTLSNVSCLSN 60  
Qy 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSLSEASITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSLSEASITKALLTRCPACSDNF 120  
Qy 121 VNLHCHNTCSFNOSLFINVTRVAQLGAGQLPVAVYEAFFYQHSFAEQSYDSCSRVRPAA 180  
Db 121 VNLHCHNTCSFNOSLFINVTRVAQLGAGQLPVAVYEAFFYQHSFAEQSYDSCSRVRPAA 180  
Qy 181 ATLAAGTMCVGSALCNQORWLNFOGDTGNGLAPLDITPHLLEPGAVSGIQLNEGV 240  
Db 181 ATLAAGTMCVGSALCNQORWLNFOGDTGNGLAPLDITPHLLEPGAVSGIQLNEGV 240  
Qy 241 ARCNEQGDVATCSQDCAASCFAIARPOALDSTFYLQMPGSLVLIITLCSVFAVVTI 300  
Db 241 ARCNEQGDVATCSQDCAASCFAIARPOALDSTFYLQMPGSLVLIITLCSVFAVVTI 300  
Qy 301 LLVGFVAPARDKSKMVDPKKGTSLSDKLSFSFTHTLGQFFQGWGTWVASWPLTILVLSV 360  
Db 301 LLVGFVAPARDKSKMVDPKKGTSLSDKLSFSFTHTLGQFFQGWGTWVASWPLTILVLSV 360  
Qy 361 IPVALLAAGLVFTELTTPVELWSAPNSQARSEKAFHDQHGFPFRFTNOVILTAPNRSY 420  
Db 361 IPVALLAAGLVFTELTTPVELWSAPNSQARSEKAFHDQHGFPFRFTNOVILTAPNRSY 420  
Qy 421 RYDSILLGPNFSGIILDLLELLELQERLRLHQLVMSPEAQNRNLSQDICVAPLNPDNT 480  
Db 421 RYDSILLGPNFSGIILDLLELLELQERLRLHQLVMSPEAQNRNLSQDICVAPLNPDNT 480  
Qy 481 SLVDDCINSILQYFQNNRTLLLTANQTLWGQTSQVDWKDHFLYCANAPLTKDGTALAL 540  
Db 481 SLVDDCINSILQYFQNNRTLLLTANQTLWGQTSQVDWKDHFLYCANAPLTKDGTALAL 540  
Qy 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600  
Db 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600  
Qy 601 RAFOREWAGNFQVTFMAERSLEDEINRTTAEDLPFATSYIVIFLYISIALGYSWSRV 660  
Db 601 RAFOREWAGNFQVTFMAERSLEDEINRTTAEDLPFATSYIVIFLYISIALGYSWSRV 660  
Qy 661 MVDKATLGLGGVAVVLGAVMAAMGFSSYLGRSSVLIVQVPPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVVLGAVMAAMGFSSYLGRSSVLIVQVPPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSEACFFLGTALTPMPAVRTFALTSGLA 780  
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSEACFFLGTALTPMPAVRTFALTSGLA 780  
Qy 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKVQDELALPKDSYLLDYFLFLNRYFEVGAPV 840  
Db 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKVQDELALPKDSYLLDYFLFLNRYFEVGAPV 840  
Qy 841 LHWITRGVVLILLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVVLILLFALFGVLSYSMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Qy 901 YFVTTILGYNFSSSEAGNNAICSSAGCANNFSTOKI QYATEFFPROSYLAIPASSWVDDFDW 960

```

Db 901 YFVTTILGYNFSSAGNNAICSSAGCNNSFTQKIQATFPPEQSYLAIPASSWVDDFIDW 960
Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCNCKMSITMGSVRPSVQFHKYLPWFNDLRP 1020
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCNCKMSITMGSVRPSVQFHKYLPWFNDLRP 1020
Qy 1021 NIKCPKGGGLAAVSTSVNLSDGQVL-----ASRPMAYH 1053
Db 1021 NIKCPKGGGLAAVSTSVNLSDGQVLDTVAILSPRLSYSGTISAHCNLYLLDASRPMAYH 1080
Qy 1054 KPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPYPTITNVFYEQVLTILPEGL 1113
Db 1081 KPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPYPTITNVFYEQVLTILPEGL 1140
Qy 1114 FMLSCLCLVPTFAVSCILLGLDLSRGLNLLSIVMILVDVGVFMALWDISYNAVSLINLVS 1173
Db 1141 FMLSCLCLVPTFAVSCILLGLDLSRGLNLLSIVMILVDVGVFMALWGISYNAVSLINLVS 1200
Qy 1174 AVGMSVEFVSHITRSPASTKPTWLERAKEATISMGSAVPAGVAMTNLPGLIIVLGLAKAQ 1233
Db 1201 AVGMSVEFVSHITRSPASTKPTWLERAKEATISMGSAVPAGVAMTNLPGLIIVLGLAKAQ 1260
Qy 1234 LIQIFPFRNLNLTLLGLHGLVFLPVLISYVGPVNPALALEQKRAEAAVAVMVASCP 1293
Db 1261 LIQIFPFRNLNLTLLGLHGLVFLPVLISYVGPVNPALALEQKRAEAAVAVMVASCP 1320
Qy 1294 NHPSRVSTADNIYVNHSPFGSIKGAAGISNFIPLNNGRQF 1332
Db 1321 NHPSRVSTADNIYVNHSPFGSIKGAAGISNFIPLNNGRQF 1359

RESULT 10
US-10-450-763-53052
; Sequence 53052, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450, 763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 53052
; LENGTH: 1344
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (901)..(917)
; OTHER INFORMATION: OESTROGEN RECEPTOR SIGNATURE domain identified by eMATRIX,
; accession number PR00543D, p-value=1.355e-09, raw score of 10.87
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1344)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-53052
```

```

Query Match 94.6%; Score 6536; DB 5; Length 1344;
Best Local Similarity 96.7%; Pred.No. 0;
Matches 1271; Conservative 7; Mismatches 21; Indels 16; Gaps 4;

Qy 1 MAEAGLRGWLWALLRLAQSPYTTIHQPGYCAFYDECGKNPELSGSIAMTLSNVSCLSN 60
Db 1 MAEAGLRGWLWALLRLAQSPYTTIHQPGYCAFYDECGKNPELSGSIAMTLSNVSCLSN 60
```

```

Qy 61 TPARKITGHLILLLOKICPRLYTGNTQACCSAKQLVSLSEASLSITKALLTRCPACSDNF 120
Db 61 TPARKITGHLILLLOKICPRLYTGNTQACCSAKQLVSLSEASLSITKALLTRCPACSDNF 120
Qy 121 VNLHCNTCSNQSLFINVTRVAQAGAGOLPAVAYEAFVQHSFABOSYDSCSRVRPAA 180
Db 121 VNLHCNTCSNQSLFINVTRVAQAGAGOLPAVAYEAFVQHSFABOSYDSCSRVRPAA 180
Qy 181 ATLAVGTMCGVYGSALCNAQRWLNFGDGTGNGLAPLDTIFHLLPEQOAVGSGIQPLNEGV 240
Db 181 ATLAVGTMCGVYGSALCNAQRWLNFGDGTGNGLAPLDTIFHLLPEQOAVGSGIQPLNEGV 240
Qy 241 ARCNSQGGDDVATCSQDCAASCP--AIAAPQALOSTFYLGQMPGSLVLIILLCSPFVAV 298
Db 241 ARCNSQGGDDVATCSQDCAASCPXPAPRPTPSTW--RQMPGSLVLIILLCSPFVAV 298
Qy 299 TILLGFRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFQGGTGWASWPLTILVL 358
Db 299 TILLGFRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFQGGTGWASWPLTILVL 358
Qy 359 SVIPVVALAAGLVFTLTTPVELWSAPNSQARSEKAFHDQHFQFPFRTNQVILTAPNRS 418
Db 359 SVIPVVALAAGLVFTLTTPVELWSAPNSQARSEKAFHDQHFQFPFRTNQVILTAPNRS 418
Qy 419 SYRYSLLIGPKNFSGILDLDLLLLELQERLHLQVMSPEAQARNISLQDICYAPLNPD 478
Db 419 SYRYSLLIGPKNFSGILDLDLLLLELQERLHLQVMSPEAQARNISLQDICYAPLNPD 478
Qy 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDG TAL 538
Db 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDG TAL 538
Qy 539 ALSCMADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGPRLPAQAALWEAFLE 598
Db 539 ALSCMADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGPRLPAQAALWEAFLE 598
Qy 599 ENRAFQRMAGMFQVTFMAERSLEDEINRTTAEDLPFATSVIVIFLYISLALGSYSSWS 658
Db 599 ENRAFQRMAGMFQVTFMAERSLEDEINRTTAEDLPFATSVIVIFLYISLALGSYSSWS 658
Qy 659 RVWDSKATLGLGGVAVVLGAVMAAGPFSYLGIRSSLVILQVVPFLVSGADNIFIFV 718
Db 659 RVWDSKATLGLGGVAVVLGAVMAAGPFSYLGIRSSLVILQVVPFLVSGADNIFIFV 718
Qy 719 LEYQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSEATCFFLGALTTPMAVTFALTSG 778
Db 719 LEYQRLPRRPGEPREHVHIGRALGRVAPSMLLCSLSEATCFFLGALTTPMAVTFALTSG 778
Qy 779 AVILDFLLQMSAFVALLSLDSKRQBSRLDVCCCKVQBELPPPGQEGLLGFFQKAYAP 838
Db 779 AVILDFLLQMSAFVALLSLDSKRQBSRLDVCCCKVQBELPPPGQEGLLGFFQKAYAP 838
Qy 839 FLLHWITRGVV-----LILFLALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLNRYF 894
Db 839 FLLHWITRGVVVPSQLLFLALFGVSLYSMCHI SVGLDQELALPKDSYLLDYFLNRYF 898
Qy 895 EVGAPVYVFTTLGVNFSSEAGNNAICSSAGCNNSFTQKIQATFPPEQSYLAIPASSWV 954
Db 895 EVGAPVYVFTTLGVNFSSEAGNNAICSSAGCNNSFTQKIQATFPPEQSYLAIPASSWV 958
Qy 955 DDFIDWLTSPSSCCRLYISGPNKDKFCPTVNSLNCNCKMSITMGSVRPSVQFHKYLPW 1014
Db 955 DDFIDWLTSPSSCCRLYISGPNKDKFCPTVNSLNCNCKMSITMGSVRPSVQFHKYLPW 1018
Qy 1015 FLNDRPNIKCPKGGGLAAVSTSVNLSDGQVLASRPMAYHKLKNSQDYTEALRAARELAA 1074
Db 1015 FLNDRPNIKCPKGGGLAAVSTSVNLSDGQVLASRPMAYHKLKNSQDYTEALRAARELAA 1078
Qy 1075 NITADLRKVPGTDPAFEPYPTITNVFYEQVLTILPEGLFMLSCLCLVPTFAVSCILLGLD 1134
Db 1075 NITADLRKVPGTDPAFEPYPTITNVFYEQVLTILPEGLFMLSCLCLVPTFAVSCILLGLD 1138
Qy 1135 LRSGLLNLLSIVMILVDVGVFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPASTK 1194
```

```
Db 1139 LRSGLNLSIWLLVDVGFMAWLGISYNAVSLINLVSAGMSVEFVSHITRSPALSTK 1198
Qy 1195 PTWLERAKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIFFRLNLLITLLGLLHG 1254
Db 1199 PTWLERAKEATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIFFRLNLLITLLGLLHG 1258
Qy 1255 LVFLPVILSYVGPVNPALALEOKRAEEA-----VAAMVAVASCPNHPRSVST 1301
Db 1259 LVFLPVILSYVGPVNPALALEOKRAEEGSGSHGGPLAQITPPESPOLTTSMST 1313

RESULT 11
US-10-621-758A-2
; Sequence 2, Application US/10621758A
; Publication No. US20040093629A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JBO1603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-621-758A-2

Query Match 78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

Qy 1 MAEAGLRGWLWALLRLAQSEPTTTHQPGYCAFYDECCKNPGLSGMLTSLNSVCLSN 60
Db 1 MAEAWL-GWLLWALLSAAQELTPKHEAGVCTFYECCKNPGLSGMLTSLNSVCLSN 59
Qy 61 TPARKITGDHLILLOKICPRLYTGPNQO-ACCSAKQLVSLSEASITIKALLTRCPACSDN 119
Db 60 TPARRVITGEHALLQRICPRLYNGPNTTACCSTKQLLSLESSITIKALLTRCPACSDN 119
Qy 120 FVNILCHNTCSPNOSLFINTVTRVQAGQLPAVVAYEAFYQHSFASQSDSCSRVRPA 179
Db 120 FVSLHCHNTCSPDQSLFINTVTRVVERGAGEPPAVVAYEAFYQSFASKAYESCQVRIPA 179
Qy 180 AATLAVGTMCGVGSALCNAORWLNFGDGTGNGLAPLIDITFHLLEPGOAVSGTQPLNEG 239
Db 180 AASLAVGSMCGVGSALCNAORWLNFGDGTGNGLAPLIDITFHLLEPGOALPDGQIPLNGK 239
Qy 240 VARNESQGDVATCSCQDCACPAIARPOALDSTFYLQMPGSLVLIILICSVFAVVT 299
Db 240 IAPCNESQGDSDSAVCSQDCACAPVIPPPEALRPSFYMGRMPCWALIIITFAVFLLS 299
Qy 300 ILLVGFVRVAPARDKSKVDPKKGTSLSDKLSPSTHTLLGQPFQGWGTWASWPLTILVLS 359
Db 300 AVLRLRVWNRNKNKAEQEPKPLPHKHLSPHTILGRFFQWGTWASWPLTILVLS 359
Qy 360 VIPVVALAAGLVFTELTDPVELWSAPNSQARSKAFHDQHFQFPFRTNQVILTAPNRSS 419
Db 360 FIVVIALAAGLTFTELTDPVELWSAPNSQARSKAFHDQHFQFPFRTNQVILTAPNRSS 419
Qy 420 YRYDSLLGPNKFGIILDLLELLELQERLRLQWSPSAQRNLSIQDICVAPLMPDN 479
Db 420 YKYDSLLGSKNFGSILSLOFLLELLELQERLRLQWSPSAQRNLSIQDICVAPLMPFN 479
Qy 480 TSLVDCINSILQYFQNNRTLLLTANOTLMGQTSQVDWKDHFLYCANAPLTKDGTSLA 539
Db 480 TSLVDCINSILQYFQNNRTLLLTANOTLMGQTSQVDWKDHFLYCANAPLTKDGTSLA 539
```

```
Db 480 TSLSDCCVNSLLQYFQNNRTLLMLTANQTNGQTSQVDWKDHFLYCANAPLTKDGTSLA 539
Qy 540 LSCMADYGAFVFFFLAIGGYKGKDYSEAEALIMTFSLNNTYPAGDPRIQAQKLWEEAFLEE 599
Db 540 LSCMADYGAFVFFFLAIGGYKGKDYSEAEALIMTFSLNNTYPAGDPRIQAQKLWEEAFLEE 599
Qy 600 MRAFORMAGMFOVTTAERSLEDEINRTTAEDLPATPSYIIVIFLYISALSGYSWSR 659
Db 600 MESFORNTSDKQVAFSAERSLEDEINRTTIOQLPVPFVASYIIVFVLYISALSGYSRCSR 659
Qy 660 VMYDSKATLGLGVAVVAVLGAAMGFYSYLGRSSILVILQVVPFLVLSVGADNIFLVL 719
Db 660 VAVESKATLGLGVIVVVGAVLAAMGFYSYLGPSSLVIIQVVPFLVLAAGNIFLVL 719
Qy 720 EYORLPRRPGEPREHVHIGRALGRVAPSMLLCSLSEAI CFFELGALTMPAVRTFALTSLGA 779
Db 720 EYORLPRMPGEPREAHIGRTLGSAVPSMLLCSLSEAI CFFELGALTMPAVRTFALTSLGA 779
Qy 780 VILDELLQMSAFVALLSLDSKROEASRLDVCCCKVQDELPPPGQEGGLILGFFQKAYAPF 839
Db 780 IILDELLQMTAFVALLSLDSKROEASRPDVLCCFSTRKLPKPKKGLLLRFRKIYAPF 839
Qy 840 LLHWITRGVLLLFALFGVSLYSMCHISVGLDQELALPKDSVLLDYFLFLNRYEFGAP 899
Db 840 LLHRFIRPVVMLLFLTLFGANLYLMCNINVLGDLQELALPKDSVLLDYFLFLNRYLEVGP 899
Qy 900 VYFVTTLGVNFSSEAGNMAICSSAGCNFSTFKIOYATEFPFQSYALIPASSWVDFFID 959
Db 900 VYFVTTSGFNFSSEAGNMAICSSAGCKSFSLTKIOYASEFPDQSYVAIAASSWVDFFID 959
Qy 960 WLTP-SSCRLYISGPNKDKFCPTVNSLNCNKMSTWGSVRPSVEQPHKYLPMFLND 1018
Db 960 WLTPSSCCRLYIRGPHKDEFCEPSTDTSFNCLNKNMNRILGVRPTAEQPHKYLPMFLND 1019
Qy 1019 RPNKCPKGLAAAYSTSVNLTSQGVLASRFMAYHKLKNSQDYTEALRAARELANITA 1078
Db 1020 PPNIRCPKGLAAAYRTSVNLSSDQVITASQFMAYHKLPLRNSQDTEALRASRLAANITA 1079
Qy 1079 DLKRVGTDPAFVPFYTITNVFYEQYLILPGLPMLSLCLVPTTFAVSCLLILGLDLRSG 1138
Db 1080 DLKRVGTDPNFVFPFYTISNVFYQYLTVLPEGIPTLALCFVPTFVVCVLLGLDMCSG 1139
Qy 1139 LMLLSIVMLVDVTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPALSTKPTWL 1198
Db 1140 ILNLLSILMLVDVTIGLMAVWGISYNAVSLINLVTAVGMSVEFVSHITRSPALSTKPTWL 1199
Qy 1199 ERAKATISMGSAVFAGVAMTNLPGLVLGLAKAQLIQIFFRLNLLITLLGLLHGLVFL 1258
Db 1200 ERAKATVFMGSAVFAGVAMTNFPGLIILGFAQQLIQIFFRLNLLITLLGLLHGLVFL 1259
Qy 1259 PVILSYVGPVNPALALEOKRAEEA--VAAMVAVASCPNHPRSVSTADNIYNHSPSGSI-KG 1317
Db 1260 PVILSYLGPVNPALVQVEEKLAEEA-AVAPEPCPOYPSPADADAN--VNYGAPLHAHG 1316
Qy 1318 AGAISNLFNNGRQF 1332
Db 1317 ANAARSSLPKSDQKF 1331
```

```
RESULT 12
US-10-663-208A-2
; Sequence 2, Application US/10663208A
; Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JBO1603K2 US
; CURRENT APPLICATION NUMBER: US/10/663,208A
; CURRENT FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442
```

```
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-663-208A-2

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGLRGWLWALLRLAQSSEPTTTHQPGYCAFYDECGKPELSGSLMTLSNVCSLN 60
DB 1 MAEAGLRGWLWALLRLAQSSEPTTTHQPGYCAFYDECGKPELSGSLMTLSNVCSLN 59
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTQ-ACCSAKQLVSLSEASLSITKALLTRCPACSDN 119
DB 60 TPARKITGDHLLILLOKICPRLYTGPNTTFFACCTKQLLSLESSMSITKALLTRCPACSDN 119
QY 120 FVNLHCHNTCSNQSLFINVTRVAQAGOLPAWVAEAFYQHSFAEQSYDSCSRVRPA 179
DB 120 FVNLHCHNTCSNQSLFINVTRVVERGAGEPPAVVAEAFYQHSFAEQSYDSCSRVRPA 179
QY 180 AATLAVGTCGYYGALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSIQPLNKG 239
DB 180 AATLAVGTCGYYGALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSIQPLNKG 239
QY 240 VARCNESQDDVATCSCQDCAAPARPOALDSTFYLGOMPGSLVLIILICSVFVAVT 299
DB 240 IAPCNESQDDSAVSCQDCAASCPIVPEALRPSFTYMGRLPGWLALIIITFAVFLLS 299
QY 300 ILLVGRVAPARDKGMVDPKXGTSLSDKLSFTHTLLGQFFQGMGTWVASWPLTILVLS 359
DB 300 AVLVRVSVNRKNKAEQEPAPKPLPHKHLSPHTILGRFFQNGTRVASWPLTILVLS 359
QY 360 VIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAFHDHGPFFRTNQVILTAPNRSS 419
DB 360 FIVJIALAAGTFTLTDPVELWSAPNSQARSEKAFHDEHGPFFRTNQVILTAPNRSS 419
QY 420 YRYSLLGPKNFGSILDLLELLEQLERLHLQVMSPEAORNISLQDICYAPLNPYN 479
DB 420 KYDSLLGSKNFGSILDLLELLEQLERLHLQVMSPEAORNISLQDICYAPLNPYN 479
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLGQTSQVDWKQHFLYCANAPLTFKDGTLA 539
DB 480 TSLDCCVNSLLQYFQNNRTLLLTANQTLGQTSQVDWKQHFLYCANAPLTFKDGTSIA 539
QY 540 LSCMADYGAPEPFLAIGYKGDYSEBALIMTSLNNYPAGDPRLAQAKLWEAFLEE 599
DB 540 LSCMADYGAPEPFLAIGYKGDYSEBALITFSLNNYPAGDPRMAQAKLWEAFLEE 599
QY 600 MRAFORMMAGMFOVFTFAERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSYKSR 659
DB 600 MESFORNTSKQVAFSAERSLEDEINRTIQDLFVFAVSIVIFLYISLALGYSYKSR 659
QY 660 VMVDSKATLGLGGVAVLGVANMAAGFFSYLGRSSVLQVVPFLVLSVGADNIFIVL 719
DB 660 VAVESKATLGLGGVAVLGVANMAAGFFSYLGRSSVLQVVPFLVAVGADNIFIVL 719
QY 720 EYQRLPRPGEPREVHIGRALGRVAPSMLLCSLSAICFFLGAITPMPAVRTFALTSLA 779
DB 720 EYQRLPRPGEPREVHIGRALGRVAPSMLLCSLSAICFFLGAITPMPAVRTFALTSLA 779
QY 780 VILDFLLQMSAFVALLSDSKQEAERLDVCCVVKPQELPPQCEGILLGFFQKAYAPF 839
DB 780 IILDFFLQMTAFVALLSDSKQEAERLDVCCVVKPQELPPQCEGILLGFFQKAYAPF 839

840 LLHWITRGVLLLFLLFALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAP 899
DB 840 LLHRFIRPVMMLLFLTLFGANLYLMCNINVLQDELALPKDSYLLDYFLFLNRYFEVGAP 899
QY 900 VYFVTTGLYNSFSEAGMNAICSSAGCNFESFTQKIQYATEFFEQSYLAIPASSWVDDFID 959
DB 900 VYFVTTGPNFSEAGMNAICSSAGCNFESFTQKIQYATEFFEQSYLAIPASSWVDDFID 959
QY 960 WLTP-SSCCRLYISGPNKDKFCPSTVNSLNCNLKNCMSITWGSVRPSEVQFHKYLPWFLND 1018
DB 960 WLTPSSCCRLYIRGPHKDEFCPTDTSFNCNLKNCNRTLGPVRPTAEQFHKYLPWFLND 1019
QY 1019 RNNICPKGGLAAYSTSVNLTSDGVLASRFMAYHKPLKNSQDYTEALRAARELANITA 1078
DB 1020 PNNIRCPKGLAAYRTSVNLTSDGVLASRFMAYHKPLKNSQDYTEALRAARELANITA 1079
QY 1079 DLRKVPGETDPAPEVEPYTITNVFYEQYLTILPEGLPMLSLCLVPTFAVSCLLGLDLRSG 1138
DB 1080 DLRKVPGETDPAPEVEPYTITNVFYEQYLTILPEGLPMLSLCLVPTFAVSCLLGLDLRSG 1139
QY 1139 LNLLSIVMILVDTVGFMAWDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWL 1198
DB 1140 ILNLISILMILVDTTGLMAVWGISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWL 1199
QY 1199 ERAKATISMGSAVFAGVAMTNLPGILVLGLAKAQLIQIFFRLMLLITLLGLLHGLVFL 1258
DB 1200 ERAKATVFMGSAVFAGVAMTNLPGILVLGLAKAQLIQIFFRLMLLITLLGLLHGLVFL 1259
QY 1259 PVILSYGDDVNPALALQKRAEBAVAVMVAACPNHPSRVSTADNIYNHSEFESI-KG 1317
DB 1260 PVVLSYLGDDVNPALALQKRAEBAVAVMVAACPNHPSRVSTADNIYNHSEFESI-KG 1317
QY 1318 AGAISNFLPNNRGOF 1332
DB 1317 ANAARSSLPKSDQKF 1331

RESULT 13
US-10-646-301A-2
; Sequence 2, Application US/10646301A
; Publication No. US20040137467A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC111 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-KI-US
; CURRENT APPLICATION NUMBER: US/10/646,301A
; CURRENT FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-646-301A-2

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGLRGWLWALLRLAQSSEPTTTHQPGYCAFYDECGKPELSGSLMTLSNVCSLN 60
DB 1 MAEAGLRGWLWALLRLAQSSEPTTTHQPGYCAFYDECGKPELSGSLMTLSNVCSLN 59
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTQ-ACCSAKQLVSLSEASLSITKALLTRCPACSDN 119
DB 60 TPARKITGDHLLILLOKICPRLYTGPNTTFFACCTKQLLSLESSMSITKALLTRCPACSDN 119
QY 120 FVNLHCHNTCSNQSLFINVTRVAQAGOLPAWVAEAFYQHSFAEQSYDSCSRVRPA 179
DB 120 FVNLHCHNTCSNQSLFINVTRVVERGAGEPPAVVAEAFYQHSFAEQSYDSCSRVRPA 179
QY 180 AATLAVGTCGYYGALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSIQPLNKG 239
DB 180 AATLAVGTCGYYGALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSIQPLNKG 239
QY 240 VARCNESQDDVATCSCQDCAAPARPOALDSTFYLGOMPGSLVLIILICSVFVAVT 299
DB 240 IAPCNESQDDSAVSCQDCAASCPIVPEALRPSFTYMGRLPGWLALIIITFAVFLLS 299
QY 300 ILLVGRVAPARDKGMVDPKXGTSLSDKLSFTHTLLGQFFQGMGTWVASWPLTILVLS 359
DB 300 AVLVRVSVNRKNKAEQEPAPKPLPHKHLSPHTILGRFFQNGTRVASWPLTILVLS 359
QY 360 VIPVVALAAGLVFTLTDPVELWSAPNSQARSEKAFHDHGPFFRTNQVILTAPNRSS 419
DB 360 FIVJIALAAGTFTLTDPVELWSAPNSQARSEKAFHDEHGPFFRTNQVILTAPNRSS 419
QY 420 YRYSLLGPKNFGSILDLLELLEQLERLHLQVMSPEAORNISLQDICYAPLNPYN 479
DB 420 KYDSLLGSKNFGSILDLLELLEQLERLHLQVMSPEAORNISLQDICYAPLNPYN 479
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANQTLGQTSQVDWKQHFLYCANAPLTFKDGTLA 539
DB 480 TSLDCCVNSLLQYFQNNRTLLLTANQTLGQTSQVDWKQHFLYCANAPLTFKDGTSIA 539
QY 540 LSCMADYGAPEPFLAIGYKGDYSEBALIMTSLNNYPAGDPRLAQAKLWEAFLEE 599
DB 540 LSCMADYGAPEPFLAIGYKGDYSEBALITFSLNNYPAGDPRMAQAKLWEAFLEE 599
QY 600 MRAFORMMAGMFOVFTFAERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSYKSR 659
DB 600 MESFORNTSKQVAFSAERSLEDEINRTIQDLFVFAVSIVIFLYISLALGYSYKSR 659
QY 660 VMVDSKATLGLGGVAVLGVANMAAGFFSYLGRSSVLQVVPFLVLSVGADNIFIVL 719
DB 660 VAVESKATLGLGGVAVLGVANMAAGFFSYLGRSSVLQVVPFLVAVGADNIFIVL 719
QY 720 EYQRLPRPGEPREVHIGRALGRVAPSMLLCSLSAICFFLGAITPMPAVRTFALTSLA 779
DB 720 EYQRLPRPGEPREVHIGRALGRVAPSMLLCSLSAICFFLGAITPMPAVRTFALTSLA 779
QY 780 VILDFLLQMSAFVALLSDSKQEAERLDVCCVVKPQELPPQCEGILLGFFQKAYAPF 839
DB 780 IILDFFLQMTAFVALLSDSKQEAERLDVCCVVKPQELPPQCEGILLGFFQKAYAPF 839
```

```
QY 120 FVNHLCHNTCSNOSLFINTVTRVAQLGAGQLPAVVAYEAFYQHSFAQSDYSCSRVPA 179
Db 120 FVSLHCHNTCSNOSLFINTVTRVVERGAGEPPAVVAYEAFYQHSFAQSDYSCSRVPA 179
QY 180 AATLAVGTCMGVYGSALCNAQRWLNFOQDTGNGLAPLDITPHLLPEQAVGSGIQPLNEG 239
Db 180 AASLAVGSMCGVYGSALCNAQRWLNFOQDTGNGLAPLDITPHLLPEQAVGSGIQPLNEG 239
QY 240 VARNESQDDVATCSODCAASCPAIPALDSTFYLQMPGSLVLIILCISVPAVVT 299
Db 240 IAPCNESQDDSAVCSODCAASCPVIPPPEALRPSFYMGMPGWLALIIIFTAVFVLLS 299
QY 300 ILLVGRVAPARDKSKMVDPKKTSLSDKLSFSTHTLLQGFQCGWGTWASWPLTILVLS 359
Db 300 AVLVLRRVVRNKNKAEGQAEAPKLPKHKLSPHTILGRFFQNGWTRVASWPLTILVLS 359
QY 360 VIPVVALAAGLVTELTTPDVELWSAPNSQARSEKAFHQHFGPFFRTNQVILTAPNRSS 419
Db 360 FIVVIALAAGLVTELTTPDVELWSAPNSQARSEKAFHQHFGPFFRTNQVILTAPNRSS 419
QY 420 YRYDSLLGPKNFGSILDLLELLELQERLRLHQLVWSPQAQRNLSLQDICYAPLNDN 479
Db 420 YKYDSLLGSKNFGSILDLLELLELQERLRLHQLVWSPQAQRNLSLQDICYAPLNDN 479
QY 480 TSLYDCCINSLLQYFQNNRTLLLTANOTLMGQTSOVDWKDHFYCANAPLTFKDGTLA 539
Db 480 TSLDCCVNSLLQYFQNNRTLLLTANOTLMGQTSOVDWKDHFYCANAPLTFKDGTLA 539
QY 540 LSCMADYGAPVFPFLAIGYKGYSEABALIMTFLNNYPAGDPRLAQAKLWEAEFL 599
Db 540 LSCMADYGAPVFPFLAIGYKGYSEABALIMTFLNNYPAGDPRLAQAKLWEAEFL 599
QY 600 MRAFORMAGQVTFPHERSLEBDEINRTTAEUPLIPATSVIVIFLYISLALGYSWSR 659
Db 600 MESFORNTSDKQFAVFAESRLEBDEINRTTQDLFPVAVSYIIVFLYISLALGYSRCSR 659
QY 660 VMVDSKATLGGVAVVLGAVMAAGPFSYLGIRSSVLQVPPFLVLSVGADNIFLVL 719
Db 660 VAVESKATLGGVIVVLGAVLAAMGFYSLGVFSSVLIIQVFFVLVAVGADNIFLVL 719
QY 720 EYQRLPRPGPREVHIGRALGRVAPSMLLCSLSEALCFILGALTMPAVRTFALTSLGA 779
Db 720 EYQRLPRPGPREVHIGRALGRVAPSMLLCSLSEALCFILGALTMPAVRTFALTSLGA 779
QY 780 VILDFLLQMSAFVALLSDSKROEASRLDVCCKVPQELPPPGQGGELLILGFKAYAPF 839
Db 780 IILDPLQMTAFVALLSDSKROEASRPDLVLCFSTRKLPPLPKKEGILLFFPKIYAPF 839
QY 840 LLHWITRGVLLLFALFGVSLYSWCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGP 899
Db 840 LLHFRFIRPVWMLLFLEFANLYLNCINVLGDELALPKDSYLLDYFLFLNRYFEVGP 899
QY 900 VYFVTTLGYNFSSAGMNAICSSAGCNNSFSTQKIYATEPPEQSYLEIPAASSWDDFID 959
Db 900 VYFVTTCFNFSSEAGMNAICSSAGCKSFSLTKIYASEFPDQSYVAIAASSWDDFID 959
QY 960 WLTP-SSCCRILYISGPNKDKCPSTVNSLNCIKCMSTMGSVRPSVEQFHLYLPWFLND 1018
Db 960 WLTPSSCCRILYIRGPHKDEFCPSDTSTFNCLKNQCMNRTLGPVRPTAEQFHLYLPWFLND 1019
QY 1019 RPNLCKPKGGLAAYSTS VNLTS DGOVLASRFMAVHKPLKNSQDYTEALRAARELANITA 1078
Db 1020 PPNLCKPKGGLAAYSTS VNLTS DGOVLASRFMAVHKPLKNSQDYTEALRAARELANITA 1079
QY 1079 DLKRVPGTDPAFEVPPYITNVFYEQYLITLPEGLFMLSCLVPTFAVSCILLGLDLRS 1138
Db 1080 DLKRVPGTDPAFEVPPYITNVFYEQYLITLPEGLFMLSCLVPTFAVSCILLGLDLRS 1139
QY 1139 LNLLSIYVILVDTVGFMAWDISYNAVSLINLSAVCMGSVEFVSHITRSFAISTKPTWL 1198
Db 1140 LNLLSIYVILVDTVGFMAWDISYNAVSLINLSAVCMGSVEFVSHITRSFAISTKPTWL 1199

RESULT 14
US-10-736-769-2
; Sequence 2, Application US/10736769
; Publication No. US20040161838A1
; GENERAL INFORMATION:
; APPLICANT: Altmann, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC111 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K3-US
; CURRENT APPLICATION NUMBER: US/10736,769
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 10/663,208
; PRIOR FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-736-769-2

Query Match 78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAAAGL-GWLLWALLSAAQGLYTPKHEAGVCTFYEECGKNPELSGGLTSLSNVCLSN 59
Db 1 MAAAGL-GWLLWALLSAAQGLYTPKHEAGVCTFYEECGKNPELSGGLTSLSNVCLSN 59
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTQ-ACCSAKQIVLSLEASISITKALLTRCPACSDN 119
Db 60 TPARKITGDHLLILLOKICPRLYTGPNTQ-ACCSAKQIVLSLEASISITKALLTRCPACSDN 119
QY 120 FVNHLCHNTCSNOSLFINTVTRVAQLGAGQLPAVVAYEAFYQHSFAQSDYSCSRVPA 179
Db 120 FVSLHCHNTCSNOSLFINTVTRVVERGAGEPPAVVAYEAFYQHSFAQSDYSCSRVPA 179
QY 180 AATLAVGTCMGVYGSALCNAQRWLNFOQDTGNGLAPLDITPHLLPEQAVGSGIQPLNEG 239
Db 180 AASLAVGSMCGVYGSALCNAQRWLNFOQDTGNGLAPLDITPHLLPEQAVGSGIQPLNEG 239
QY 240 VARNESQDDVATCSODCAASCPAIPALDSTFYLQMPGSLVLIILCISVPAVVT 299
Db 240 IAPCNESQDDSAVCSODCAASCPVIPPPEALRPSFYMGMPGWLALIIIFTAVFVLLS 299
QY 300 ILLVGRVAPARDKSKMVDPKKTSLSDKLSFSTHTLLQGFQCGWGTWASWPLTILVLS 359
Db 300 AVLVLRRVVRNKNKAEGQAEAPKLPKHKLSPHTILGRFFQNGWTRVASWPLTILVLS 359
QY 360 VIPVVALAAGLVTELTTPDVELWSAPNSQARSEKAFHQHFGPFFRTNQVILTAPNRSS 419
Db 360 FIVVIALAAGLVTELTTPDVELWSAPNSQARSEKAFHQHFGPFFRTNQVILTAPNRSS 419
```

[illegible]

RESULT 15

US-10-621-758A-12

US 10 021 758A 12  
; Sequence 12, Application US/10621758A

; Publication No. US20040093629A1

; GENERAL INFORMATION:

; APPLICANT: Altmann, Scott W

APPLICANT: Wang, Luquan

```

; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1l1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JE01603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1333
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-621-758A-12

```

Query Match	78.3%	Score 5407	DB 4	Length 1333
Best Local Similarity	77.2%	Pred. No. 0		
Matches 1028	Conservative 126	Mismatches 174	Indels 4	Gaps 4
Qy	4	AGLRGMLLWALLRLAQSPPYTIHQHGVYCAFDECGKNPGLSGSLMTLSNVSCLSNTPA	63	
Db	3	AAMQGWLLWALLLNSAQGLYPTTHKAGCTFYEECGKNPGLSGGLTSLSNVSCLSNTPA	62	
Qy	64	RKITGDHLLILLOKICPRLYTGNP-TOACCSAKQLVSLSEASLSITKALLTRCPACSNFVN	122	
Db	63	RHVTGDHALLQRCVCPRLYNGPNDTYACCSKQLVSLDSSLSITKALLTRCPACSNFVS	122	
Qy	123	LHCNHTCSNOSLFINVTRVAQLGAGLPANVAYEAFYQHSFAEQSYDSCSRVVRPAAT	182	
Db	123	IHCNHTCSPDQSUFINVTRVQRDPQLPAVAYEAFYQRSFABKAYESCSRVRIPAAAS	182	
Qy	183	LAVGTMCGVYGVSALCNAQRWLNFGQDTGNGLAPLDITFFHLLPFGQAVGSGIOPLNEGVAR	242	
Db	183	LAVGSMCGVYGVSALCNAQRWLNFGQDTGNGLAPLDITFFHLLPFGQALADGMKPLDGKI	242	
Qy	243	CNESQGDVATCSCQDCAASCAPAIARPAQLDSTFYLGQMPGSLVLIILCSVPVAVVTILL	302	
Db	243	CNESQGEDSAACSQCQDCAASCVPVPPPPALRPSFYMGMRPGMLALIIITFAVFLVLSVVL	302	
Qy	303	VGFPRVAPARDKSNWDPKKGTSLSDKLSFTHTLLGQFFQGGTGWASWPLTITLVLVSP	362	
Db	303	VYLRVASNRKNKTAGSCQAPNLPKRRSPHTVYLGRRFESNGTRVASWPLVLVLA	362	
Qy	363	VWLAAGLVFTELTDPVWELGAPNSQARSEKAFHQHFGFPFRTNQVLTAPNRSRYR	422	
Db	363	VIALSVGLTFIELTDPVWELGAPKSAQKAKAFHDEHFGFPFRTNQIVPTAKNRSYKY	422	
Qy	423	DSLGLGPKNFGSLDLDLALLELEQERLRHLQVWSPEAQRNLSIODICYAPLNPDNTSL	482	
Db	423	DSLGLGPKNFGSLDLLOELLEQERLRHLQVWSHEAQRNLSIODICYAPLNPHNTSL	482	
Qy	483	YDCINSLSLQYFONNRTLLLTANQTLMGQTSQVQWKDHFLYCANAPLTFFKQGTALALSC	542	
Db	483	TDCVNSLLQYFONNHTLLLLTANQTLNGQTSLVMDKQHFLYCANAPLYTKDGTALALSC	542	
Qy	543	MADYGAPVPFFLAIQYKQKQYSEBALIMTFSLNYPAGDPRRLAQAKLWBEAFLEEMRA	602	
Db	543	IADYGAPVPFFLAVGQYQGTYSAEALITPFSINNYPADDDPRMAHAKLWBEAFLEKEMOS	602	
Qy	603	FORRMAGMQVFTFAERSLEDEINRTAEDLPFATSYIVIFYLSALGVSYSMSRVVW	662	
Db	603	FQRSTADKQIAPFAERSLEDEINRTIQDLFPFALISYILVFLYLSALGVSYSRVRVAV	662	
Qy	663	DSKATLGLGGVAVLGAVMMAAGPFSYLGIRSSLVILQVWPELVLSVAGDNIFFVLRYQ	722	
Db	663	DSKATLGLGGVAVLGAVVAAGMFGYSLGVPSLSVIIQVWPELVLAGADNIFI	722	
Qy	723	RLPRRGEPREVHIQEALGRVAPSMLLCSLSBAICFFLGAITPMPAVTFTALTSGLAVIL	782	
Db	723	RLPRMPEQGEAREAHIGRTLSGVAPSMLLCSLSBAICFFLGAITSPMPAVTFTALTSGLATIF	782	
Qy	783	DFLLOWSAPVALLSLSDKREQASRLDVCCKVQBELPPPGQEGELLLGPFQKAYAPFLI	842	

Db	783	DFLLQMTAFVALLSLDSKROEASRPDVVCCFSSRNLPFPKQKGLLCPFRKTYTFFPLH	842
Qy	843	WITRGVLLFLALPGVSLYSMCHISVGLDOELAKPKDSLYDYFLFLNRYFVSGAPVYF	902
Db	843	RFIRPVVLLFLVLFGANLJLMCNISVGLDOOLAKPKDSLYDYFLFLNRYLEVGPPVYF	902
Qy	903	VTTILGVNFSSEAGNNAICSSAGCNPFFSTOKIOVATEFFPQSYLAIAPASSWVDDFIDWLT	962
Db	903	DTTSGYNFSTEAGNNAICSSAGCESFBLTOKIOVASEFFNQSYVAIAASSWVDDFIDWLT	962
Qy	963	P-SSCCRLXYISGPNKKFCFSTVNSLCLNCKMSITWGSVRPSVEQPHKYLPMFLNDRPN	1021
Db	963	PSSSCCRIYTRGPHKDFCFSTDTSFNCLKNCMNRTLGPVRPTTEQPHKYLPMFLNDTPN	1022
Qy	1022	IKCPKGLAAYSTSVNLTSGQVULASRFMAYHKLKNSQDYTEALRAARELANIITADLR	1081
Db	1023	IRCPKGLAAYRTSVNLTSSDQIIITASQFMAYHKLKNSQDFTALRASRIILANITAE LR	1082
Qy	1082	KVPGTDPAFEVFPYTIITNVFYOQLTTLPEGLFMLSICLVPTFAVSCLLGLDLRSLGN	1141
Db	1083	KVPGTDPNFEVFPYTIISNVFYQQYTLVLPGEIFTALCFPFTFVVCYLLGLDLIRSGILN	1142
Qy	1142	LLSTVMTLLVDTVGPMALWDISYNVAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER A	1201
Db	1143	LLSLIIMTLLVDITGUMVAVGSLSYNAVSLINLVTAVGMSVEFVSHITRSFAVSTRKTR LER A	1202
Qy	1202	KEATISMGSAVFAGVAMTNIPGIIIVLVLGLAKAQLTIQIFFRNLNLTLLGLLHGLVFLPVI	1261
Db	1203	KDATIFMSAVFAGVAMTNIPGIIILGLFAQAQLTIQIFFRNLNLTLLGLLHGLVFLPVV	1262
Qy	1262	LSYVGPDPVNPALALEQKRAEEAAVMVAVGCPNHPRSVSTADNTIYNVHSPEGS- IKGAGA	1320
Db	1263	LSYLGPDVNQALVLEEXLATEA-AMVSEPSCPQPPFPADANTSQYVNVYGNPFPIPEINA	1321
Qy	1321	ISNFLPNNGRQF 1332	
Db	1322	ASSSLPKSDOKF 1333	

Search completed: April 7, 2006, 19:37:11  
Job time : 194 secs



November 2005

Published\_Applications Nucleic Acid and Published\_Applications Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases: older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New). Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

**THIS PAGE IS BLANK**



Qy	301	LLVGRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQPFQGWGTWASWPLTILVLSV	360
Db	301	LLVGRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQPFQGWGTWASWPLTILVLSV	360
Qy	361	IPVVALAAGLVFTELTTDPVELSAPNSQARSKAHFDHQHFGFFRTNVLTPAKRSSY	420
Db	361	IPVVALAAGLVFTELTTDPVELSAPNSQARSKAHFDHQHFGFFRTNVLTPAKRSSY	420
Qy	421	RYDSLILGPKNFSGILDLLELLELLEQLERHLQVMSPEAQRNLSIQDICYAPLNPNDT	480
Db	421	RYDSLILGPKNFSGILDLLELLELLEQLERHLQVMSPEAQRNLSIQDICYAPLNPNDT	480
Qy	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHPLYCANAPLTFKDGOTAL	540
Db	481	SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHPLYCANAPLTFKDGOTAL	540
Qy	541	SCMADYGAPVPPFLAIGGYGKDYSEABALIMTFSLNYPAGDPRLAQAQKLWBEAFLEEM	600
Db	541	SCMADYGAPVPPFLAIGGYGKDYSEABALIMTFSLNYPAGDPRLAQAQKLWBEAFLEEM	600
Qy	601	RAFORRMAGMPQVFTTAERSLEDEINRTAEDLPFATSVIVIFLYISLALGSYSMSRV	660
Db	601	RAFORRMAGMPQVFTTAERSLEDEINRTAEDLPFATSVIVIFLYISLALGSYSMSRV	660
Qy	661	MVDSKATLGLGGVAVLGVANMAAGFFSYLGIRSLVILQVVPFLVLSVGADNIFIFVLE	720
Db	661	MVDSKATLGLGGVAVLGVANMAAGFFSYLGIRSLVILQVVPFLVLSVGADNIFIFVLE	720
Qy	721	YORLPRRGEPREVHIGALGRVAPSMLLCSLSBAICFFLGALTPMPAVRTFALTSLGLAV	780
Db	721	YORLPRRGEPREVHIGALGRVAPSMLLCSLSBAICFFLGALTPMPAVRTFALTSLGLAV	780
Qy	781	ILDFFLQMSAFVALLSLDSKQASRLDVCCKVQPELPPPGQEGELLGPFQKAYAPFL	840
Db	781	ILDFFLQMSAFVALLSLDSKQASRLDVCCKVQPELPPPGQEGELLGPFQKAYAPFL	840
Qy	841	LHWITRGVLLFLFALFGVLSYMSCHISVGLDQELALPKDSYLDYDFLFLNRYFEVGPV	900
Db	841	LHWITRGVLLFLFALFGVLSYMSCHISVGLDQELALPKDSYLDYDFLFLNRYFEVGPV	900
Qy	901	YFVTTLGYNFSSEAGMNAICSSAGCNFSTFKIQIYATEFPFQSYLAIPASSWDDFDW	960
Db	901	YFVTTLGYNFSSEAGMNAICSSAGCNFSTFKIQIYATEFPFQSYLAIPASSWDDFDW	960
Qy	961	LTPSSCCRLYISGPNKDFCPTVNSLNCNMSITMGSVRPSVEQFHKLPMFLNDRP	1020
Db	961	LTPSSCCRLYISGPNKDFCPTVNSLNCNMSITMGSVRPSVEQFHKLPMFLNDRP	1020
Qy	1021	NIKCPKGLAAYSTSVNLTSQVLAASFMAVHKPLKNSQDYTEALRAARELANITADL	1080
Db	1021	NIKCPKGLAAYSTSVNLTSQVLAASFMAVHKPLKNSQDYTEALRAARELANITADL	1080
Qy	1081	RKVPGTDPAFVFPYTTINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGLL	1140
Db	1081	RKVPGTDPAFVFPYTTINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGLL	1140
Qy	1141	NLLSIVMILVDTVGFMAWMDISYNAVSLINLVSAVGMSEVFVSHITRSPASTKPTWLER	1200
Db	1141	NLLSIVMILVDTVGFMAWMDISYNAVSLINLVSAVGMSEVFVSHITRSPASTKPTWLER	1200
Qy	1201	AKEATISMGSAPVAGVAMTNLCIILVLGAKAQLIQIPFRLNLLITLGLLHGLVFLPV	1260
Db	1201	AKEATISMGSAPVAGVAMTNLCIILVLGAKAQLIQIPFRLNLLITLGLLHGLVFLPV	1260
Qy	1261	ILSYGPDVNPALALQKAEAEAVAAVMVASCPNHPHSRVSTADNIYVNHSPFGS1KGACA	1320
Db	1261	ILSYGPDVNPALALQKAEAEAVAAVMVASCPNHPHSRVSTADNIYVNHSPFGS1KGACA	1320
Qy	1321	ISNFLPNNGRQF 1332	
Db	1321	ISNFLPNNGRQF 1332	

RESULT 2

US-10-995-561-952

Sequence 952, Application US/10995561

Publication No. US2005027054A1

GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.

TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF DETECTION AND USES THEREOF

TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF DETECTION AND USES THEREOF

TITLE OF INVENTION: DETECTION AND USES THEREOF

FILE REFERENCE: CL001559

CURRENT APPLICATION NUMBER: US/10/995,561

CURRENT FILING DATE: 2004-11-24

NUMBER OF SEQ ID NOS: 85702

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 952

LENGTH: 1278

TYPE: PRT

ORGANISM: Homo sapiens

US-10-995-561-952

Query Match 34.8%; Score 2402.5; DB 6; Length 1278;

Best Local Similarity 39.8%; Pred. No. 2.6e-181;

Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

Qy 7 RGWLLWALLRLAQSEPTTTHQPGYCAFYDECG-----KNPELSSGLMTLSNVCSLS 59

Db 4 RGLALGLLLLLCPAQVFSQ-----SCVWYEGCIAYGDKRYNCEYSG-----46

Qy 60 NTPARKITGDHLLILLOKICPRLYTGNTQACCSAKOLVSEASLSITKALLTRCPACSDN 119

Db 47 --PPKPLPKDGYDLVQELCPGFFFG-NVSLCCDVRQLQTLKONLQPLQLFGRCPSCFTN 103

Qy 120 FVNHLCHNTCSNQSLSFINVTR----VAOLGAGQLPAVVAYEAFYQHSFAEOSYDSCSRV 175

Db 104 LMLFCELTCSRQSOFLNVTATEDYVDPVTNQTKNVKELYVYVQSQSFANAMYNACRDV 163

Qy 176 RVPAAATLAVGTWCMGYVGSALCNAQRWLNFGQGTGNGLAPLDIT-----FHLLPQGA 228

Db 164 EAPSSNDKALGLLCKGQADA-CNATNWEYFMKNQGOAPFTTTPVFSDFPVH-----215

Qy 229 VSGGIQPLNEGVARNCSQGDVATCSQDCAASCPAIAAPQ-----ALDSTFYL 279

Db 216 ---GMEPMNATKGCDESDEVTAPCSCQDCSIVCGPKQPPPPPPAPWTILGDMYVIM 272

Qy 280 QMPGSLVLIILGCVFAVVTILL-----VGFVAPARDKSKMVDPKKGTSL 326

Db 273 WITYMAELLVFFGAFVAVWCYKRYFVSEYTPIDSNIAFSV-NASDKGE-----ASCC 324

Qy 327 DKLSFSTHTLLGQPFQGWGTWASWPLTILVSVIPVVALAAGLVFTELTTDPVELWSAP 386

Db 325 DPVSAAFEGCLRLRFTWGSFCVRNPGCVIFFSLVFTACSSGLVFRVYRTNTPVDLWSAP 384

Qy 387 NSQARSKAFHDQHFQGFPEFTNQVILTPAKRSSYRYSLSLLGPK-NFSGILDLDLLELL 445

Db 385 SSOARLSEKYPDQHFQGFPEFTNQVILTPAKRSSYRYSLSLLGPK-NFSGILDLDLLELL 444

Qy 446 ELQERLRLHLOWSPAQRNLSIQDICYAPLNPNTSYDCCINSLLQYFQNNRTLLLLTA 505

Db 445 DLQTAIEN--ITASVDNETVTLQDICLAPLSPYNT--NCTILSVLNYFQNSHSLVDHKK 499

Qy 506 NQTLMGQTSQVDWKDHPLYCANAPLTFKDGOTALALSCHADYGAPEFPLAIGGYGKDY 565

Db 500 GDDFF--VVADYTHFELYCVRAPASLNDTSLHDPCLTGTFGGVFPVWLVGGYDDQNYN 556

Qy 566 EAEALIMTFSLNYPAGDPRLAQAQKLWBEAFLEEMRAFORRMAGMPQVFTTAERSLEDEI 625

Db 557 NATALVITFPVNNYNDTEKLQRAQAKWEKFINFKYKN---PNLITSTFAERSIEDEL 613

Qy 626 NRTTAEDLPFATSVIVIFLYISLALGSYSMSRWMSVDKATLGLGGVAVLGVAMAAMG 685

Db 614 NRESDSLVFTVVISYAIMFLYISLALGHKICRRLVDSKSLGILAGILVLSVACSLG 673



Db 489 PHSLONNIRTSLEKRRGSGVPWASRPEATLQDFPGTVVYISKVQCHPA--VYRLSLNA 546  
QY 911 SSEAGNVAICSSAG--CNWF-----SFTQKIQVATEPEPSYLAIPASSWVDDFDWL 961  
Db 547 SLFAPMQAVSPGDGEVPSQVYRAPFGNFTKKLTACMTVGLLQAASPSRKNW-----L 600  
QY 962 TPSSC-----CRLYIS--GP 974  
Db 601 TTLACDAKRGKWFDSFYVATKEQHQTRKLYFAQSHKPPFHGRVCMAPPGCLSSSPDGP 660  
QY 975 NKDKFCPTVNSLNCNKMSITMG-----SVRPSVEQFHKYLPMF-----LN- 1017  
Db 661 TKGFF---FVPSKVPKARLSATFGFNPVCVNTGCGKPAVRPLVDTGAMVVFVGLIYNR 717  
QY 1018 ----DRPNIKCPKGLAAYSTSVNLTS-----GQVL 1045  
Db 718 TRQVDNHVIGDP--GSVVYDSSFDLFKEIGHLCHLCKAIAANSELVVKPGQAQCLPSGYSI 775  
QY 1046 ASRFMAYHKPLKNSOD-----YTEALRAARELAANITADLRKVPDPAPEVF-P 1094  
Db 776 SSFLOMLHPECKELPEPNLLPGQLSHGAVGVRGVRQWISMAFESTTYKGS-SFQTYSD 834  
QY 1095 YTTINVFYEQYLTILPEG-----LFM-----LSLCL--VP 1122  
Db 835 YLRWESFLOQLOALPEGSVLRRGFQTCHEMKQIFMEIVGVQSAQCLGLVLSLICVAAVA 894  
QY 1123 TFAVSCLLGLDLRSLGLNLLSIVMLVDTVCFMALWDISYNAVSLINLVSAVGMSEVFV 1182  
Db 895 VFTTHILL-----LLPVLSILGIVCLV--TIMYWSGEM--GAVEAISLSILVSSVDYC 947  
QY 1183 SHITRSFAIS-----TKPTWLERAKEATISMGSAVFAGVAMTNLPGI----- 1224  
Db 948 VHLVEGYLLAGENLPPHQAEDARTQW--RTLEAVRHVGVAIVSSALTTVIATVPLFFC 1005  
QY 1225 LVGLAKAQLOIQIFFRNLTLTLLGLHGLVFLPVILSVYGVDPVNPALALEQKRAE--E 1282  
Db 1006 IIAAPFAK-----FGKIVALNTGVSILYTLTVSTALLGIMAPS-----SFTRTTSFLK 1053  
QY 1283 AVAAMVVA 1290  
Db 1054 ALGAVILLA 1061

## RESULT 5

US-11-242-459-40  
; Sequence 40, Application US/11242459  
; Publication No. US20060035835A1  
; GENERAL INFORMATION:  
; APPLICANT: TANIYAMA, Yoshio  
; APPLICANT: KITA, Shunbun  
; APPLICANT: SATOMI, Tomoko Komiya  
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof  
; FILE REFERENCE: 2703USOP  
; CURRENT APPLICATION NUMBER: US/11/242,459  
; CURRENT FILING DATE: 2005-10-03  
; PRIOR APPLICATION NUMBER: US/10/239,316  
; PRIOR FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: PCT/JP01/02279  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: JP2000-088595  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 59  
; SEQ ID NO 40  
; LENGTH: 1392  
; TYPE: PRT  
; ORGANISM: Human  
US-11-242-459-40

Query Match 2.8%; Score 195.5; DB 7; Length 1392;  
Best Local Similarity 19.5%; Pred. No. 8.6e-07;  
Matches 200; Conservative 129; Mismatches 320; Indels 379; Gaps 47;

QY 576 LNNYPAGDPLRAQAQKLWEAFLEEMRAFORMAGHFQVTTAERSLED----- 623

Db 405 LPNTYISVDDR-----WEE-----ORAKFQSFVTVYVYVAMLAKQSTKSVQVLYGSTD 449  
QY 624 ----EINRTAED--LPIFATSYIVIFLYISLALGSSYSSSRVMDVKATLGLGG-VAVV 676  
Db 450 LFDYEVARTFNMDMLAFISSCIAALVYILTSCSVLSPFGI-----ASIGLSCLVALF 504  
QY 677 LGAVMAAMGFSSYIGIRSSVILQVVPFLVLSVGADNIFIVLEYORLPR--RPGEPREHV 735  
Db 505 LYHVVFGI---QYLGIILNG---VAAFVIVGIVDDVDFVINTYRQATHLEDQPLRMH 556  
QY 736 IGRALGRVAPSMLLCSLSEALCFPLGALTMPAVRTALTSLGLAVI---LDFFLOMSAFV 792  
Db 557 TVQTAGK---ATFTSTLTTAAVAAVFSQIPAVHDFGLFMSLIVSCCWLAVLTMPAAL 613  
QY 793 ALLSLDSKROEASRLDVC--CCVKPQBLPPPGQ----- 823  
Db 614 GLMSLYLAPLESSCOTSCONCSRKTSLHPFDVATPEQVGGSPAQGPPIPYLDDDDIPLL 673  
QY 824 -----GEGLLGFFQKAYAP-----FLHWHI-----TRGVVL 850  
Db 674 EVBEEPVSLELGDVSLVSVSPGLOPASNTGSRGHLIVQLQELLHHWVLSAVKSRWVIV 733  
QY 851 LLFLALFGVSL-----YSM-----CHISVGLDOE- 874  
Db 734 GLFVSIILSLVFPASRLRPARAPLLFRPDTNIOVLDLKYNLSAEGISCTCSGLFQEK 793  
QY 875 -----LALPK-----DSYLLDY--FLFLNRYFEVGPVYVVTTLGYNF 910  
Db 794 PHSLONNIRTSLEKRRGSGVPWASRPEATLQDFPGTVVYISKVQCHPA--VYRLSLNA 851  
QY 911 SSEAGNVAICSSAG--CNWF-----SFTQKIQVATEPEPSYLAIPASSWVDDFDWL 961  
Db 852 SLFAPMQAVSPGDGEVPSQVYRAPFGNFTKKLTACMTVGLLQAASPSRKNW-----L 905  
QY 962 TPSSC-----CRLYIS--GP 974  
Db 906 TTLACDAKRGKWFDSFYVATKEQHQTRKLYFAQSHKPPFHGRVCMAPPGCLSSSPDGP 965  
QY 975 NKDKFCPTVNSLNCNKMSITMG-----SVRPSVEQFHKYLPMF-----LN- 1017  
Db 966 TKGFF---FVPSKVPKARLSATFGFNPVCVNTGCGKPAVRPLVDTGAMVVFVGLIYNR 1022  
QY 1018 ----DRPNIKCPKGLAAYSTSVNLTS-----GQVL 1045  
Db 1023 TRQVDNHVIGDP--GSVVYDSSFDLFKEIGHLCHLCKAIAANSELVVKPGQAQCLPSGYSI 1080  
QY 1046 ASRFMAYHKPLKNSOD-----YTEALRAARELAANITADLRKVPDPAPEVF-P 1094  
Db 1081 SSFLOMLHPECKELPEPNLLPGQLSHGAVGVRGVRQWISMAFESTTYKGS-SFQTYSD 1139  
QY 1095 YTTINVFYEQYLTILPEG-----LFM-----LSLCL--VP 1122  
Db 1140 YLRWESFLOQLOALPEGSVLRRGFQTCHEMKQIFMEIVGVQSAQCLGLVLSLICVAAVA 1199  
QY 1123 TFAVSCLLGLDLRSLGLNLLSIVMLVDTVCFMALWDISYNAVSLINLVSAVGMSEVFV 1182  
Db 1200 VFTTHILL-----LLPVLSILGIVCLV--TIMYWSGEM--GAVEAISLSILVSSVDYC 1252  
QY 1183 SHITRSFAIS-----TKPTWLERAKEATISMGSAVFAGVAMTNLPGI----- 1224  
Db 1253 VHLVEGYLLAGENLPPHQAEDARTQW--RTLEAVRHVGVAIVSSALTTVIATVPLFFC 1310  
QY 1225 LVGLAKAQLOIQIFFRNLTLTLLGLHGLVFLPVILSVYGVDPVNPALALEQKRAE--E 1282  
Db 1311 IIAAPFAK-----FGKIVALNTGVSILYTLTVSTALLGIMAPS-----SFTRTTSFLK 1358  
QY 1283 AVAAMVVA 1290  
Db 1359 ALGAVILLA 1366

RESULT 6

```
US-11-096-568A-6778
; Sequence 6778, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6778
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(127)
; OTHER INFORMATION: Ceres Seq. ID no. 15168324
US-11-096-568A-6778

Query Match          2.8%; Score 191; DB 7; Length 127;
Best Local Similarity 44.3%; Pred. No. 6.4e-08;
Matches 35; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

QY 1199 ERAKEATISGSAVFAGVAMTNLPGILVLGLAKAQLIQIFPRNLNLTLLGLLHGLVFL 1258
DB 34 QRAKTALCTMGASVFSGITLTKLVGLVLCFSTSIQIFVYFQMYLALVLGLHGLVFL 93

QY 1259 PVILSYGPDVNPALALEQ 1277
DB 94 FVVLISLFGPLRYTVIKEQ 112

RESULT 7
US-11-096-568A-6779
; Sequence 6779, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6779
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(85)
; OTHER INFORMATION: Ceres Seq. ID no. 15168325
US-11-096-568A-6779

Query Match          2.5%; Score 173; DB 7; Length 85;
Best Local Similarity 44.3%; Pred. No. 9.6e-07;
Matches 31; Conservative 19; Mismatches 20; Indels 0; Gaps 0;

QY 1208 MGSVAFGAVAMTNLPGILVLGLAKAQLIQIFPRNLNLTLLGLLHGLVFLPVILSYGVP 1267
DB 1 MGSVAFSGITLTKLVGLVLCFSTSIQIFVYFQMYLALVLGLHGLVFLPVILSLFGP 60

QY 1268 DYNPALALEQ 1277
DB 61 PLRYTVIKEQ 70

RESULT 8
US-11-242-459-35
; Sequence 35, Application US/11242459
; Publication No. US20060035835A1
```

```
; GENERAL INFORMATION:
; APPLICANT: TANIYAMA, Yoshio
; APPLICANT: KITA, Shunbun
; APPLICANT: SATOMI, Tomoko Komiyama
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/11/242,459
; CURRENT FILING DATE: 2005-10-03
; PRIOR APPLICATION NUMBER: US/10/239,316
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 35
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Human
US-11-242-459-35

Query Match          2.1%; Score 143; DB 7; Length 445;
Best Local Similarity 25.4%; Pred. No. 0.0024;
Matches 69; Conservative 41; Mismatches 106; Indels 56; Gaps 12;

QY 576 LNNYPAGDPRLAQAQKLWEAFLEEMRAFORRMAGMFQVTFTAERSLED----- 623
DB 100 LPNYSVDDR-----WEE-----QRAKEFSFVTVYVAMLAQSKVQVLYGGTD 144

QY 624 ----EINRTABD--LPIFATSYVIFLYISIALGYSYSSWRVMDSKATLGLGG-VAVV 676
DB 145 LFDYEVRRTFNNDMLIAFISSCIAALVILTSCSVLFSFFGI-----ASIGLSCLVALF 199

QY 677 LGAVMAAMGFPSYLGIRSSLVILQVVPFLVLSVGADNIFVLEYQRLPR-RGEPREVH 735
DB 200 LYHVVEGI---QYLGIING-----VAAFVIVGIVDDVFVINTYRQATHLEDPPQURMIH 251

QY 736 IGRALGRVAPSMLLCSLSEAICFFLGALTPMPAVRTFALTSGIAVI---LDFFLLQMSAFV 792
DB 252 TVQTAGK---ATFFSTLTAAAYANVFSQIPAVHDFGLFMSLIVSCCWLAVLVTPPAL 308

QY 793 ALLSLDSKQERASRLDVC--CCVKPQELPPPG 822
DB 309 GLWSLYLAPLESSCOTSCHQNGSRKTSLHFPG 340

RESULT 9
US-11-242-459-34
; Sequence 34, Application US/11242459
; Publication No. US20060035835A1
; GENERAL INFORMATION:
; APPLICANT: TANIYAMA, Yoshio
; APPLICANT: KITA, Shunbun
; APPLICANT: SATOMI, Tomoko Komiyama
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/11/242,459
; CURRENT FILING DATE: 2005-10-03
; PRIOR APPLICATION NUMBER: US/10/239,316
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 34
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Human
US-11-242-459-34

Query Match          2.1%; Score 143; DB 7; Length 456;
Best Local Similarity 25.4%; Pred. No. 0.0025;
```





Db 329 TTPVVSASIH-EVVKTLGE-----AIIIVFLVMFLQNF-----RATLIPTI 370  
Qy 671 GGVAIVLG--AVNAAMGFFSYLGIRSLVILQVVPFLVSVG--ADNIFIVLEYORLPR 726  
Db 371 AVPVLLGTGFLVLAAPGF-----SINTLTWFG-WVLAIGLLVDDAIVVVENVERVMA 421  
Qy 727 RPG-EPREV---HIGRALGR-VAPSMLLCSLSAICFFFLGALTTPMPAVRTFALTSLGLAVI 781  
Db 422 EELSPREARKSMGOIQGALVGIAMVLSAVFLPMAFFGGSTGVI--YRQFSITIVSAMA 479  
Qy 782 LDFLLQMSAFVALLSLDSKQEAASRLDVCCC--VKPOELPPPGQEGLLGPFQKAYAP 838  
Db 480 L-----SVIVALI-----LTPALCATMLAPIEKGDGHEKHGFFGFWNFRNLS 522  
Qy 839 FLHWTIRGVV-----LLFLALFGVSLYSMCHISVGL-----DQ-----ELALPKD 880  
Db 523 -TTHGYERGVASILKRAPYLLIYVIVVAGMIWMFTRIPTAPLPDEDQGVLFPAQVQTPPG 581  
Qy 881 S-----YLLDYFLFLNRYFEVGAIVYFVTTLTGYNFSSE-----913  
Db 582 SSAERTQVVVDSMRREYLLEK-----BSSSVSSVFTVTGFNFAGRGQSSGMAFTMLXP 633  
Qy 914 ----AGMAI---CSSAGCNPFSTQKIQYATEPEQSVLAIPASSWDDDFD-----959  
Db 634 WEERPGSENSVELAKRAQHFFSKDAMVFA--PAPPSVLELGNATGDFLQDOQGVG 691  
Qy 960 ----WLTSSCCRLYISGPNKDFPCPTVN-----SLNCLKNCMSI 996  
Db 692 HEVLLQARNKFLMLAAQNPAQRVRPNGMSDEPOYKLEIDDEKASALGVSLADINSTVSI 751  
Qy 997 TWGSV-----RPSVEQFHKLWFLNDRPN-----IKCPKGGLA---AYST 1034  
Db 752 AWGSSVYNDFIDRGRVKRV-----YLOGRPARMPDLSKWMYVRNDRKGEMVPPFNAFAT 805  
Qy 1035 ----SVNLTSQGVLASFPWAYHKPLKNSODYTEALRAARELANIITADLRKVPCTDP 1088  
Db 806 GKWEYGPCLERYNGVPAMEILGEPAPGLSSGD---AMAAVEE---IVKQLPKGVG---855  
Qy 1089 AFEVFPTTITVVFYEQLT-----ILPEGLFMLSIC-----LVPTFAVS 1127  
Db 856 ----YSWTGLSYEERLSSGQAPALYALSLLVVFLCLAALYESWSIPFSVMLVPLGVI 909  
Qy 1128 CULLGLDLSGLNLLSIVMILVDVTVGFMALWDISNVAVSLNL-----V 1172  
Db 910 GALLATSMR-GLSNDVFFVQGLLTITGLSA-----KNALLIVEFAKELHEQKGIVEAAI 963  
Qy 1173 SAVGMSVEFVSHITRSFALSTKPTWLERAKEATISGSAVFAGVAMTNLPGILVL-GLAK 1231  
Db 964 EACRMRLRPVMTSLAFILGVVP-----LAISTG-----AGSGSQHAIGTGVIGGMVT 1011  
Qy 1232 AOLIQIIFPRL 1242  
Db 1012 ATVLAIFWVPL 1022

RESULT 12

US-10-467-657-2648

; Sequence 2648, Application US/10467657

; Publication No. US20050260581A1

; GENERAL INFORMATION:

; APPLICANT: CHIRON SPA

; APPLICANT: FONTANA Maria Rita

; APPLICANT: PIZZA Mariagrazia

; APPLICANT: MASIGNANI Vega

; APPLICANT: MONACI Elisabetta

; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS

; FILE REFERENCE:

; CURRENT APPLICATION NUMBER: US/10/467,657

; PRIOR FILING DATE: 2003-08-11

; PRIOR APPLICATION NUMBER: GB-0103424.8

; PRIOR FILING DATE: 2001-02-12

; NUMBER OF SEQ ID NOS: 9218

; SOFTWARE: Seqwin99, version 1.04

; SEQ ID NO 2648

; LENGTH: 1067

; TYPE: PRT

; ORGANISM: Neisseria gonorrhoeae

US-10-467-657-2648

Query Match

Best Local Similarity 19.4%; Score 134; DB 6; Length 1067;

Matches 185; Conservative 139; Mismatches 354; Indels 274; Gaps 44;

Qy 452 RHLOW-SPEAQR--NISLQDICYAPLNPDPNTSLDYDCCINSLLQYQNNRTLLLLLTANQT 508

Db 180 RAMRIVWDPKLQNTNLSFADVGS-LSAQNQISAGSIGSLPAV--RGQT---VTVATVT 233

Qy 509 LMGQTSQVDWKDHFYCANAPLTFKQGTALALSCMADYGAPVFPFLAIGYKGDYSEAE 568

Db 234 AQQLGTAEEFGNVILRANT-----DGSNIYLKDAKVGLGM-----EDVSSST 277

Qy 569 ALI-----MTFSLNNYPAGDPRLAQAKLWEBAFLBEMRAFORRMAGMFQVTFABRSLE 622

Db 278 RLNGVNTTCMAVMLSN--SGN-AMATAKAVKERLAVLEKYFPQGMS--WKTPTDTSKFVE 332

Qy 623 DEINRTAEDLIPATSIYIVILYISLALGSSYSSHSRVMVDSKATLGLGGAIVVLGAVMA 682

Db 333 ISIEKVIHTLIBAMVLVFMVYFLQNI--RYTLPTIIVVPISL---LGGFA-----379

Qy 683 AMGFYSYLGIRSLVILQVVPFLVLSVGADNIFIFVLEYORLPRRPGEPREHVHIGRALGR 742

Db 380 ---FISYMG-SINVLTMFAMILVIGIVVDDAIVVENVERIMAGEGLPPEKATKKAMCQ 435

Qy 743 VAPSM-----LCISLSEAICFFLGALTTPMPAVRTFALTSLGLAVILDLLQMSAFVALLSL 797

Db 436 ISGAVIGITAVLISVFPVPLAMPFSGAAGNI--YKQFALTWASSI-----AFSAFLAL---484

Qy 798 DSKREASRLDVCCCVKQDELPPPGQEGE--LLGFFQKAYAPF-----LHWNIT 845

Db 485 -----TLTPALCATMLKTIPKGHEEKKGFFGFWNKKFDSWTHGYEGRVAKVLKRTF 536

Qy 846 RGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAIVYFVIT 905

Db 537 RMVVYIGLAVGV-----FLPWR-----LPTSFLPT 563

Qy 906 LGYNFSSEAGNNAICSSAGC-----NNFSTQKIQYATEPPE-QSYLAIPASSWDD-----956

Db 564 EDQGFV---MVSQVLPAGATKERTDATLAQVTLAKSIPETENIITVSGFSFGSQGNM 619

Qy 957 -----FIDW--LTSSCCRLYIS-----GPNKDFPCPTVN-----S 986

Db 620 AMGFALDKDWNERTAGSDAVAVAGKLTGMMMTLKDQGFISVPPPILELNGSGLSIN 679

Qy 987 LNCLKNCMSITMGSVRPSVEQPHKYLWPF-----LNDRENIKCPKGGGLAAVYSTVN 1037

Db 680 LQDRNNTGHTALLAKRNELIQWRASGLFDPSTVTRAGGLEDSFQIKIDINRAAAAGGIS 739

Qy 1038 LTSQGVLASRF-MAYHKPLKNSODYTEALRAARELANIITADLRK-----VP---1084

Db 740 PADIRALTALASLSSSYSDFPNQGLQRVMVQADEARMQPADILNLTVPNKSAGVAVPLS 799

Qy 1085 -----GTD-----PAFEVPPYITVNFVEQYL-----1106

Db 800 TIATVSWENGTEQSVRFNGYPSMKLSASPATGVSTGOAMAAVQKMVDLGGYSGFEGWGQ 859

Qy 1107 -----TILPEGLFMLSICLV-----PTFAVSLCLLGL-----DLR 1136

Db 860 SREBAKGGSQTLILYGLAVAVFLVLAALYESWSIPLAVILVPLGLIGAAAGVTGRNLF 919

Qy 1137 SGLL-NLLSIVMILVDTVGFMALWDIS--YNVAVSLINLV-----SAVGSVE-----F 1181

Db 920 EGLGSPSPANDIYFQGVFTVMGLSAKNAIILIEPAKDLQAQGSKSAVEAAEARLRF 979

Qy 1182 VSHITRSPA--ISTKPTWLERAKEATI--SHGSAVFAGVAMTNLPGILVLGL 1229

Db 980 RPIIMTSPAFILGVVPLYIAAGASSASQRAIGTTTFWGMVLVGTLLSVFLVPL 1031



```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2020
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2020

Query Match      1.8%; Score 126; DB 6; Length 801;
Best Local Similarity 19.9%; Pred.No.0.12;
Matches 105; Conservative 85; Mismatches 180; Indels 157; Gaps 25;

QY 841 LHWITR-----GVVLLFLALFGVS-----LYSM-----CHISVGLDQ 873
Db 62 LNWMPHIGMNFNLYVDGLGFLSLITIGISLVVLYSIGVLSKSEQLGNFYCYLLLFMGA 121
QY 874 ELALP-KDSYLDY-----FLFNRYFEVGAPVY-----FVTTLGYNPSEAGM 916
Db 122 MLGVLSDNFIILYLFWELTSFSSFLISFWREKKASIYCAQKSLIITVLG-GLSMLGGI 180
QY 917 NAICSSAGCNFSFTQKIQYATEPPEQSYL-----AIPASSWVDDFIDW 960
Db 181 --ILLSLATDTFSIQAMISKASDIQNSPPFILVMILFMIGAFTKSAQVPFYIWLDPAMEA 238
QY 961 LTPSSC-----CRLYISGPNKDKFCPS-----TVNSLNLCKNCMSITMGSVRPSVE 1006
Db 239 PTFVSAYLHSATWVKAGLYLIARITPIFAISEGWVWIT-----LVGLITLFWASLNATKQ 294
QY 1007 QFHKYLFWFLNDRPNIKCPKGLAAYSTSVNLTSDGQVLASRFMAYHKPLKNSQDYTEAL 1066
Db 295 HDLK-----GILAPSTVSQLGIMSMGLGICAVSYHYQGANSQLYVAGF 337
QY 1067 RAARELAANI-----TADLRKVPGTDPAPFEVPEPYTITNVFYEQYLT 1107
Db 338 VAAIFHLINHATFKGALFMITGDIHSTGTDRVKKLGG---LTTIMPISFT-----LT 387
QY 1108 ILPEGLFMLSCLVPTFAVSCLLLGLDLRSGLNLNLSIVMLVDTVGFPMALWDISYNAVS 1167
Db 388 VIT-----TLSMAGVPPP--NGFLSKEKFLESINVTNLMSLNTLG-----ILLPIIA 435
QY 1168 LINLVSAGMSVEFVSHITRSPAIKPTWL-ERAKBATISMGSAVFAGVAMTNLPGLIV 1226
Db 436 IIGSIPTFVYSIKFILHI---PFGSYKPEALPKQAHESSILM--LISPIILTSL--VIV 487
QY 1227 LGLAKAQLIQIFFFFRNLNLTLLG-----LLHGLVFLPVILSYVG 1266
Db 488 FGLFPISILTQSIIEPASVAVSQTSNITAEFHLFHGIT--PAFLSTIG 532
```

Search completed: April 7, 2006, 19:37:45  
Job time : 33 secs

**THIS PAGE IS BLANK**

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpn** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapn** and **.rapn**

***Because they contain data that is confidential, the results of Pending database searches should not be left in the case .***

**THIS PAGE IS BLANK**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:22:20 ; Search time 569 Seconds  
(without alignments)  
3235.086 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 MAEAGLRGWLWALLRLAQ.....GSIKAGAGATSNFLPNNGRQF 1332

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 7861189 seqs, 1381955077 residues

Total number of hits satisfying chosen parameters: 7861189

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_AA\_Main.\*  
1: /cgn2\_6/ptodata/1/paa/US066 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US073 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US074 COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US075 COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US076 COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US077 COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US078 COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US079 COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US080 COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US081 COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US082 COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US083 COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US084 COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US085 COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US086 COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US087 COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US088 COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US089 COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US090 COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US091 COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US092 COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US093 COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US094 COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US095 COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US096 COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US097 COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US098 COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US099 COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US100 COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US101 COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US102 COMB.pep.\*  
32: /cgn2\_6/ptodata/1/paa/US103 COMB.pep.\*  
33: /cgn2\_6/ptodata/1/paa/US104 COMB.pep.\*  
34: /cgn2\_6/ptodata/1/paa/US105 COMB.pep.\*  
35: /cgn2\_6/ptodata/1/paa/US106 COMB.pep.\*  
36: /cgn2\_6/ptodata/1/paa/US107 COMB.pep.\*  
37: /cgn2\_6/ptodata/1/paa/US108 COMB.pep.\*  
38: /cgn2\_6/ptodata/1/paa/US109 COMB.pep.\*  
39: /cgn2\_6/ptodata/1/paa/US110 COMB.pep.\*  
40: /cgn2\_6/ptodata/1/paa/US111 COMB.pep.\*  
41: /cgn2\_6/ptodata/1/paa/US112 COMB.pep.\*  
42: /cgn2\_6/ptodata/1/paa/US113 COMB.pep.\*  
43: /cgn2\_6/ptodata/1/paa/US114 COMB.pep.\*

44: /cgn2\_6/ptodata/1/paa/US600 COMB.pep.\*  
45: /cgn2\_6/ptodata/1/paa/US601 COMB.pep.\*  
46: /cgn2\_6/ptodata/1/paa/US602 COMB.pep.\*  
47: /cgn2\_6/ptodata/1/paa/US603 COMB.pep.\*  
48: /cgn2\_6/ptodata/1/paa/US604 COMB.pep.\*  
49: /cgn2\_6/ptodata/1/paa/US605 COMB.pep.\*  
50: /cgn2\_6/ptodata/1/paa/US606 COMB.pep.\*  
51: /cgn2\_6/ptodata/1/paa/US607 COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6909	100.0	1332	1	PCT-US03-22467-4	Sequence 4, Appli
2	6909	100.0	1332	1	PCT-US03-40113-4	Sequence 4, Appli
3	6909	100.0	1332	1	PCT-US05-01469-4	Sequence 4, Appli
4	6909	100.0	1332	36	US-10-621-758A-4	Sequence 4, Appli
5	6909	100.0	1332	36	US-10-646-301A-4	Sequence 4, Appli
6	6909	100.0	1332	36	US-10-663-208A-4	Sequence 4, Appli
7	6909	100.0	1332	37	US-10-736-769-4	Sequence 4, Appli
8	6909	100.0	1332	37	US-10-750-386-4	Sequence 4, Appli
9	6896	99.8	1332	1	PCT-US01-04098A-1831	Sequence 1831, Ap
10	6896	99.8	1332	32	US-10-239-316-9	Sequence 9, Appli
11	6896	99.8	1332	32	US-10-258-899A-1831	Sequence 1831, Ap
12	6896	99.8	1332	32	US-10-293-244-1831	Sequence 1831, Ap
13	6896	99.8	1332	42	US-11-218-141-1831	Sequence 1831, Ap
14	6896	99.8	1332	42	US-11-242-459-9	Sequence 9, Appli
15	6872.5	99.5	1359	1	PCT-US01-04098A-1830	Sequence 1830, Ap
16	6872.5	99.5	1359	1	PCT-US03-22467-44	Sequence 44, Appli
17	6872.5	99.5	1359	1	PCT-US03-40113-44	Sequence 44, Appli
18	6872.5	99.5	1359	1	PCT-US05-01469-44	Sequence 44, Appli
19	6872.5	99.5	1359	1	PCT-US05-27579-22	Sequence 22, Appli
20	6872.5	99.5	1359	32	US-10-258-899A-1830	Sequence 1830, Ap
21	6872.5	99.5	1359	32	US-10-293-244-1830	Sequence 44, Appli
22	6872.5	99.5	1359	36	US-10-621-758A-44	Sequence 44, Appli
23	6872.5	99.5	1359	36	US-10-646-301A-44	Sequence 44, Appli
24	6872.5	99.5	1359	36	US-10-663-208A-44	Sequence 44, Appli
25	6872.5	99.5	1359	37	US-10-736-769-44	Sequence 44, Appli
26	6872.5	99.5	1359	37	US-10-750-386-44	Sequence 44, Appli
27	6872.5	99.5	1359	42	US-11-218-141-1830	Sequence 1830, Ap
28	6868.5	99.4	1359	31	US-10-170-205E-21686	Sequence 21686, A
29	6868.5	99.4	1359	48	US-60-453-050-8956	Sequence 8956, Ap
30	6868.5	99.4	1359	48	US-60-466-412-8956	Sequence 8956, Ap
31	6868.5	99.4	1344	1	PCT-US01-08631-53052	Sequence 53052, A
32	6536	94.6	1344	34	US-10-450-763-53052	Sequence 53052, A
33	6536	94.6	1344	1	PCT-US03-22467-2	Sequence 2, Appli
34	5421.5	78.5	1331	1	PCT-US03-40113-2	Sequence 2, Appli
35	5421.5	78.5	1331	1	PCT-US05-01469-2	Sequence 2, Appli
36	5421.5	78.5	1331	1	PCT-US05-27579-2	Sequence 2, Appli
37	5421.5	78.5	1331	36	US-10-621-758A-2	Sequence 2, Appli
38	5421.5	78.5	1331	36	US-10-646-301A-2	Sequence 2, Appli
39	5421.5	78.5	1331	36	US-10-663-208A-2	Sequence 2, Appli
40	5421.5	78.5	1331	37	US-10-736-769-2	Sequence 2, Appli
41	5421.5	78.5	1331	37	US-10-750-386-2	Sequence 2, Appli
42	5407	78.3	1333	1	PCT-US03-22467-12	Sequence 12, Appli
43	5407	78.3	1333	1	PCT-US03-40113-12	Sequence 12, Appli
44	5407	78.3	1333	1	PCT-US05-01469-12	Sequence 12, Appli
45	5407	78.3	1333	1	PCT-US05-27579-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
PCT-US03-22467-4  
; Sequence 4, Application PC/TUS0322467  
; GENERAL INFORMATION:  
; APPLICANT: Schering Corporation

```
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K-WI
; CURRENT APPLICATION NUMBER: PCT/US03/22467
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-22467-4

Query Match      100.0%; Score 6909; DB 1; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLGRWLLWALLRLAQAQSEPTTTHQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60
DB 1 MAEAGLGRWLLWALLRLAQAQSEPTTTHQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60
QY 61 TPARKITGDHLLILQKICPRLYTGNTQACCSAKQLVSLASLITKALLTRCPACSDNF 120
DB 61 TPARKITGDHLLILQKICPRLYTGNTQACCSAKQLVSLASLITKALLTRCPACSDNF 120
QY 121 VNLHCHNTCSNQSLFINVTRVAOLGAGOLPAVVAEAFYOHSAFQSDYSCSRVRPAA 180
DB 121 VNLHCHNTCSNQSLFINVTRVAOLGAGOLPAVVAEAFYOHSAFQSDYSCSRVRPAA 180
QY 181 ATLAGVTMCGVYGSAALCAQWLNFGQDTGNGLAPLDITFHLLEPFGQAVGSGIQPLNEGV 240
DB 181 ATLAGVTMCGVYGSAALCAQWLNFGQDTGNGLAPLDITFHLLEPFGQAVGSGIQPLNEGV 240
QY 241 ARCHESQDDVATSCQDCAACPAIARPOALDSTFYLGQMPGSLVLIILICSFVAVVTI 300
DB 241 ARCHESQDDVATSCQDCAACPAIARPOALDSTFYLGQMPGSLVLIILICSFVAVVTI 300
QY 301 LLVGRVAPADKSKWDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVWASWPLTILVLSV 360
DB 301 LLVGRVAPADKSKWDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVWASWPLTILVLSV 360
QY 361 IPVVALAAGLVFTELTDPVELWSAPNSQARSEKAFHQHGFPPRTNQVILTAPNRSSY 420
DB 361 IPVVALAAGLVFTELTDPVELWSAPNSQARSEKAFHQHGFPPRTNQVILTAPNRSSY 420
QY 421 RYDSLILGPKNPGSLDLDLLELLELQERLHLQVWSPQARNISLODICVAPLNPNT 480
DB 421 RYDSLILGPKNPGSLDLDLLELLELQERLHLQVWSPQARNISLODICVAPLNPNT 480
QY 481 SLYDCCINSLLQYFONNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKGTALAL 540
DB 481 SLYDCCINSLLQYFONNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKGTALAL 540
QY 541 SCWADYGAPVFPFLAIGYKGRDYSEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
DB 541 SCWADYGAPVFPFLAIGYKGRDYSEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
QY 601 RAFQRMAGMFQVTTAERSLEDEINRTTAEDLPFATSYIVIFLYISALGSYSSWSRV 660
DB 601 RAFQRMAGMFQVTTAERSLEDEINRTTAEDLPFATSYIVIFLYISALGSYSSWSRV 660
QY 661 MYDSKATILGLGVAVVLGAVMAAGFFSYLGRSSILVILQVVPFLVLSVGADNIFIVLE 720
DB 661 MYDSKATILGLGVAVVLGAVMAAGFFSYLGRSSILVILQVVPFLVLSVGADNIFIVLE 720
QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSEACFFLGLALTMPAVRTFALTSGLA 780
DB 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSLSEACFFLGLALTMPAVRTFALTSGLA 780
QY 781 ILDFLLQMSAFVALLSLDSKQREASRLDVCCCKVQKPELPPOGGEGLLILGFQKAYAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSKQREASRLDVCCCKVQKPELPPOGGEGLLILGFQKAYAPFL 840
QY 841 LHWITRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900
DB 841 LHWITRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900
QY 901 YFVTTILGYNFSSEAGNNAICSSAGCNNFSTQKIQYATEPEPQSYLAIPASSWVDDPIDW 960
DB 901 YFVTTILGYNFSSEAGNNAICSSAGCNNFSTQKIQYATEPEPQSYLAIPASSWVDDPIDW 960
QY 961 LTPSSCCRLYISGPNKDKFCPSVNSLNCIKNCMSITMGSVRPSVEQFHYKLPWFNDRP 1020
DB 961 LTPSSCCRLYISGPNKDKFCPSVNSLNCIKNCMSITMGSVRPSVEQFHYKLPWFNDRP 1020
QY 1021 NIKCPKGLAAYSTVNLTSQVTLASRFMAYHKPLKNSQDYTEALRAARELANITADL 1080
DB 1021 NIKCPKGLAAYSTVNLTSQVTLASRFMAYHKPLKNSQDYTEALRAARELANITADL 1080
QY 1081 RKYPGTDPAPEFPPYITINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140
DB 1081 RKYPGTDPAPEFPPYITINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140
QY 1141 NLLSIWMLVDTVGFMAWMDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200
DB 1141 NLLSIWMLVDTVGFMAWMDISYNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200
QY 1201 AKKATISMGSAVPAVAMTNLPGILVGLAKAOLIOIOPPRMLNLLITLLGLHGLVPLPV 1260
DB 1201 AKKATISMGSAVPAVAMTNLPGILVGLAKAOLIOIOPPRMLNLLITLLGLHGLVPLPV 1260
QY 1261 ILSVVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSPGSIKAGA 1320
DB 1261 ILSVVGPDVNPALAEQKRAEAAVAAVMVASCPNHPRSVSTADNIYVNHSPGSIKAGA 1320
QY 1321 ISNPLPNNGRQF 1332
DB 1321 ISNPLPNNGRQF 1332

RESULT 2
PCT-US03-40113-4
; Sequence 4, Application PC/TUS0340113
; GENERAL INFORMATION:
; APPLICANT: Schering Corporation
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K3-WI
; CURRENT APPLICATION NUMBER: PCT/US03/40113
; CURRENT FILING DATE: 2003-12-16
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2003-08-22
; PRIOR FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-40113-4

Query Match      100.0%; Score 6909; DB 1; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLGRWLLWALLRLAQAQSEPTTTHQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60
DB 1 MAEAGLGRWLLWALLRLAQAQSEPTTTHQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60
QY 61 TPARKITGDHLLILQKICPRLYTGNTQACCSAKQLVSLASLITKALLTRCPACSDNF 120
DB 61 TPARKITGDHLLILQKICPRLYTGNTQACCSAKQLVSLASLITKALLTRCPACSDNF 120
```



QY 121 VNLHCHTCSNQSLFINTVTRVAQLGAGQLPAVVAEAFYQHSFAEQSDSCSRVRVPA 180  
DB 121 VNLHCHTCSNQSLFINTVTRVAQLGAGQLPAVVAEAFYQHSFAEQSDSCSRVRVPA 180  
QY 181 ATLAVGTCMGVYGSALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV 240  
DB 181 ATLAVGTCMGVYGSALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV 240  
QY 241 ARCNEQGGDDVATCSCQDCAASCPAIPARQALDSTFYLQMPGSLVLIILCSVFAVVTI 300  
DB 241 ARCNEQGGDDVATCSCQDCAASCPAIPARQALDSTFYLQMPGSLVLIILCSVFAVVTI 300  
QY 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTIILVSV 360  
DB 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTIILVSV 360  
QY 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFQFPFRFTNQVILITAPNRSY 420  
DB 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFQFPFRFTNQVILITAPNRSY 420  
QY 421 RYDSLGLPKNFGSLDLDLLELELQERLRLHQLWSPQAQRNISLQDICYAPLNPDNT 480  
DB 421 RYDSLGLPKNFGSLDLDLLELELQERLRLHQLWSPQAQRNISLQDICYAPLNPDNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGKTALAL 540  
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGKTALAL 540  
QY 541 SCWADYGAVFPFLAIGGYKGKDYSEAEALIMTFESLNNYPAGDPRLAQAKLWEEAPLEEM 600  
DB 541 SCWADYGAVFPFLAIGGYKGKDYSEAEALIMTFESLNNYPAGDPRLAQAKLWEEAPLEEM 600  
QY 601 RAFQRMAGMFOVTTAERSLEDEINRTTAEDLPFATSYIIVIFLYISIALGYSWSRV 660  
DB 601 RAFQRMAGMFOVTTAERSLEDEINRTTAEDLPFATSYIIVIFLYISIALGYSWSRV 660  
QY 661 MYDSKATILGGAVVVLGAVMAAMGFFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLE 720  
DB 661 MYDSKATILGGAVVVLGAVMAAMGFFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLE 720  
QY 721 YORLPRRPGEPREVHIGRALGVAPSMILCSISEALCFELGALTMPAVRTTALTSGLAV 780  
DB 721 YORLPRRPGEPREVHIGRALGVAPSMILCSISEALCFELGALTMPAVRTTALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKPOELPPPGOGEGLLIGFFQKAYAPPL 840  
DB 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCKPOELPPPGOGEGLLIGFFQKAYAPPL 840  
QY 841 LHWITRGVVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
DB 841 LHWITRGVVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
QY 901 YFVTTILGNFSEAGHNAICSSAGCNPFSFTOKIQYATEFPFQSYLAIPASSWVDDFDIW 960  
DB 901 YFVTTILGNFSEAGHNAICSSAGCNPFSFTOKIQYATEFPFQSYLAIPASSWVDDFDIW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPTVNSINCLKNCMSITMGSVRPSVEQPHKILPWFINDRP 1020  
DB 961 LTPSSCCRLYISGPNKDKFCPTVNSINCLKNCMSITMGSVRPSVEQPHKILPWFINDRP 1020  
QY 1021 NIKCPKGGILAAYSTSVNLTSDQVLASRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
DB 1021 NIKCPKGGILAAYSTSVNLTSDQVLASRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080  
QY 1081 RKVPGTDPAFEPFPTITNVFYEQLTILPEGLFMLSCLVPTFPAVSCLLGLDLSRGLL 1140  
DB 1081 RKVPGTDPAFEPFPTITNVFYEQLTILPEGLFMLSCLVPTFPAVSCLLGLDLSRGLL 1140  
QY 1141 NLLSIVMILVDVTGFMALWDISVNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
DB 1141 NLLSIVMILVDVTGFMALWDISVNAVSLINLVSAGVMSVEFVSHITRSFAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQLIQIPFFRLNLLTITLGLLHGLVFLPV 1260

DB 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQLIQIPFFRLNLLTITLGLLHGLVFLPV 1260  
QY 1261 ILSYVGPDVNPALALEOKRAEBAVAAMVAVASCPNHPSPRVSTADNIYVNHSPGSIKGA 1320  
DB 1261 ILSYVGPDVNPALALEOKRAEBAVAAMVAVASCPNHPSPRVSTADNIYVNHSPGSIKGA 1320  
QY 1321 ISNFLPNNGRQF 1332  
DB 1321 ISNFLPNNGRQF 1332  
RESULT 3  
PCT-US05-01469-4  
; Sequence 4, Application PC/TUS0501469  
; GENERAL INFORMATION:  
; APPLICANT: Garcia-Calvo, Margarita  
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF IDENTIFYING LIGANDS THEREOF  
; FILE REFERENCE: 36134-PCT 074669.0111  
; CURRENT APPLICATION NUMBER: PCT/US05/01469  
; PRIOR FILING DATE: 2005-01-25  
; PRIOR APPLICATION NUMBER: US 60/537,341  
; PRIOR FILING DATE: 2004-01-16  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US05-01469-4  
Query Match 100.0%; Score 6909; DB 1; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCSLSN 60  
DB 1 MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKPELSGSLMTLSNVCSLSN 60  
QY 61 TPARKITGDHLILLQKICPRLVTGPNTQACCSAKOLVLEASLSITKALLTRCPACSDNF 120  
DB 61 TPARKITGDHLILLQKICPRLVTGPNTQACCSAKOLVLEASLSITKALLTRCPACSDNF 120  
QY 121 VNLHCHTCSNQSLFINTVTRVAQLGAGQLPAVVAEAFYQHSFAEQSDSCSRVRVPA 180  
DB 121 VNLHCHTCSNQSLFINTVTRVAQLGAGQLPAVVAEAFYQHSFAEQSDSCSRVRVPA 180  
QY 181 ATLAVGTCMGVYGSALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV 240  
DB 181 ATLAVGTCMGVYGSALCNAQRLNFGDGTGNGLAPLDITFHLLEPGQAVGSGIQPLNEGV 240  
QY 241 ARCNEQGGDDVATCSCQDCAASCPAIPARQALDSTFYLQMPGSLVLIILCSVFAVVTI 300  
DB 241 ARCNEQGGDDVATCSCQDCAASCPAIPARQALDSTFYLQMPGSLVLIILCSVFAVVTI 300  
QY 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTIILVSV 360  
DB 301 LLVGRFVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGWGTWASWPLTIILVSV 360  
QY 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFQFPFRFTNQVILITAPNRSY 420  
DB 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFQFPFRFTNQVILITAPNRSY 420  
QY 421 RYDSLGLPKNFGSLDLDLLELELQERLRLHQLWSPQAQRNISLQDICYAPLNPDNT 480  
DB 421 RYDSLGLPKNFGSLDLDLLELELQERLRLHQLWSPQAQRNISLQDICYAPLNPDNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGKTALAL 540  
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGKTALAL 540  
QY 541 SCWADYGAVFPFLAIGGYKGKDYSEAEALIMTFESLNNYPAGDPRLAQAKLWEEAPLEEM 600

Db 541 SCWADYCAPVPFFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAOKLWEEAFLEEM 600  
Qy 601 RAFORMMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSSRV 660  
Db 601 RAFORMMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSSRV 660  
Qy 661 MVDKATLGLGGVAVLGVAVLWMAAGFFSYLGRSLVILQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVLGVAVLWMAAGFFSYLGRSLVILQVVPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRPGEPREVHIGRALGRVAPSMMLCSLSAICFFLGALTPMPAVRTFALTSLGLAV 780  
Db 721 YQRLPRPGEPREVHIGRALGRVAPSMMLCSLSAICFFLGALTPMPAVRTFALTSLGLAV 780  
Qy 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCVKPQELPPPGQEGILLGFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCVKPQELPPPGQEGILLGFQKAYAPFL 840  
Qy 841 LHWITRGVVLFLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVVLFLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Qy 901 YFVTTLGVNFSSEAGMNAICSSAGCNPFSTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960  
Db 901 YFVTTLGVNFSSEAGMNAICSSAGCNPFSTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960  
Qy 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNCKMSITMGSVRPSVEQPHKYLPMFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNCKMSITMGSVRPSVEQPHKYLPMFLNDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSQGVLASRFMAYHKPLKNSQDYTEALRAARELAANIADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLTSQGVLASRFMAYHKPLKNSQDYTEALRAARELAANIADL 1080  
Qy 1081 RKVPGTDPAFVFPYTTINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFVFPYTTINVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140  
Qy 1141 NLLSVIMLVDTVGMALWDISYNAVSLINLVSAVGMSEVFSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSVIMLVDTVGMALWDISYNAVSLINLVSAVGMSEVFSHITRSFAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQILQIIEFFRLNLLITLLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAGVAMTNLPGLVLGLAKAQILQIIEFFRLNLLITLLGLLHGLVFLPV 1260  
Qy 1261 ILSYVGPDPNPALALEQKRAEEAAVAVMVASCPNHPRSRVSTADNIYVNHSPGSIKGAGA 1320  
Db 1261 ILSYVGPDPNPALALEQKRAEEAAVAVMVASCPNHPRSRVSTADNIYVNHSPGSIKGAGA 1320  
Qy 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

## RESULT 4

US-10-621-758A-4  
; Sequence 4, Application US/10621758A  
; GENERAL INFORMATION:  
; APPLICANT: Altmann, Scott W  
; APPLICANT: Wang, Luquan  
; APPLICANT: Graziano, Michael  
; APPLICANT: Murgolo, Nick  
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF  
; FILE REFERENCE: JB01603-K-US  
; CURRENT APPLICATION NUMBER: US/10/621,758A  
; CURRENT FILING DATE: 2003-07-17  
; PRIOR APPLICATION NUMBER: 60/397,442  
; PRIOR FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-621-758A-4  
Query Match 100.0%; Score 6909; DB 36; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MABAGLRGWLWALLRLAQSEPYTTHQPGYCAFYDECGKNPELSGLMTLSNVCSL 60  
Db 1 MABAGLRGWLWALLRLAQSEPYTTHQPGYCAFYDECGKNPELSGLMTLSNVCSL 60  
Qy 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Qy 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRPAA 180  
Db 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRPAA 180  
Qy 181 ATLAVGTMCGVYSALCNAORWLNFOGDTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240  
Db 181 ATLAVGTMCGVYSALCNAORWLNFOGDTGNGLAPLDITPHLLEPGQAVSGIQLNEGV 240  
Qy 241 ARCNEQGDVATCSCODCAASCAPAARPOALDSTFYLGOMPGSLVLIILCSVFAVVTI 300  
Db 241 ARCNEQGDVATCSCODCAASCAPAARPOALDSTFYLGOMPGSLVLIILCSVFAVVTI 300  
Qy 301 LLVGRVAPARDKSMVDPKKGTSLDKLSFSFTHLLGQFPQGWGTWVASWPLTILVLSV 360  
Db 301 LLVGRVAPARDKSMVDPKKGTSLDKLSFSFTHLLGQFPQGWGTWVASWPLTILVLSV 360  
Qy 361 IPVVALAAGLVFTELTTDPVELWSAPNSQARSKAHFHQHFGPFRTNOVILTAPNRSY 420  
Db 361 IPVVALAAGLVFTELTTDPVELWSAPNSQARSKAHFHQHFGPFRTNOVILTAPNRSY 420  
Qy 421 RYDSSLILGPNKFSIGLIDLDLLELLELQERLRLHQLVMSPEAQRNLSIQDICAPLNDPT 480  
Db 421 RYDSSLILGPNKFSIGLIDLDLLELLELQERLRLHQLVMSPEAQRNLSIQDICAPLNDPT 480  
Qy 481 SLVDCCLNSLLQYFQNNRTLLLTANQTLMGQTSQVDWKOHFLYCANAPLTFKDGTAIAL 540  
Db 481 SLVDCCLNSLLQYFQNNRTLLLTANQTLMGQTSQVDWKOHFLYCANAPLTFKDGTAIAL 540  
Qy 541 SCWADYCAPVPFFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAOKLWEEAFLEEM 600  
Db 541 SCWADYCAPVPFFLAIGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAOKLWEEAFLEEM 600  
Qy 601 RAFORMMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSSRV 660  
Db 601 RAFORMMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSSSRV 660  
Qy 661 MVDKATLGLGGVAVLGVAVLWMAAGFFSYLGRSLVILQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVLGVAVLWMAAGFFSYLGRSLVILQVVPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRPGEPREVHIGRALGRVAPSMMLCSLSAICFFLGALTPMPAVRTFALTSLGLAV 780  
Db 721 YQRLPRPGEPREVHIGRALGRVAPSMMLCSLSAICFFLGALTPMPAVRTFALTSLGLAV 780  
Qy 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCVKPQELPPPGQEGILLGFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCVKPQELPPPGQEGILLGFQKAYAPFL 840  
Qy 841 LHWITRGVVLFLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVVLFLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Qy 901 YFVTTLGVNFSSEAGMNAICSSAGCNPFSTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960  
Db 901 YFVTTLGVNFSSEAGMNAICSSAGCNPFSTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960  
Qy 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNCKMSITMGSVRPSVEQPHKYLPMFLNDRP 1020



Db 1321 ISNFPNNGRQF 1332

RESULT 6

US-10-663-208A-4

; Sequence 4, Application US/10663208A

; GENERAL INFORMATION:

; APPLICANT: Altmann, Scott W

; APPLICANT: Wang, Luquan

; APPLICANT: Graziano, Michael

; APPLICANT: Murgolo, Nick

; TITLE OF INVENTION: NPC111 (NPC3) AND METHODS OF USE THEREOF

; FILE REFERENCE: JB01603K2 US

; CURRENT APPLICATION NUMBER: US/10/663,208A

; CURRENT FILING DATE: 2003-09-16

; PRIOR APPLICATION NUMBER: 60/397,442

; PRIOR FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 10/621,758

; PRIOR FILING DATE: 2003-07-17

; PRIOR APPLICATION NUMBER: 10/646,301

; PRIOR FILING DATE: 2003-08-22

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 1332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-663-208A-4

Query Match 100.0%; Score 6909; DB 36; Length 1332;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLRGWLWALLRLAQSPPYTHIQPGYCAFYDECGKNPELSGLMTLSNVCSLN 60

Db 1 MAEAGLRGWLWALLRLAQSPPYTHIQPGYCAFYDECGKNPELSGLMTLSNVCSLN 60

QY 61 TPARKITGDHLLLOKICPRLVTGNTQACCSAKOLVSEASLSITKALLTRCPACSNF 120

Db 61 TPARKITGDHLLLOKICPRLVTGNTQACCSAKOLVSEASLSITKALLTRCPACSNF 120

QY 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAEQSYDCSRVRVPA 180

Db 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAEQSYDCSRVRVPA 180

QY 181 ATLAVGTWCVYGSALCNAQRWLNFGDTGNGLAPLDTTFHLLPEQAVGSGIOPLNEGV 240

Db 181 ATLAVGTWCVYGSALCNAQRWLNFGDTGNGLAPLDTTFHLLPEQAVGSGIOPLNEGV 240

QY 241 ARCNESQGDVATCSQDCAACPAIARPOALDSTFYLGOMPGSLVLIILCSVPVVTI 300

Db 241 ARCNESQGDVATCSQDCAACPAIARPOALDSTFYLGOMPGSLVLIILCSVPVVTI 300

QY 301 LLAGVRVAPARSKRWDPKKGTSLSKLSFSTHTLLGQFFQGWGTWVASMPLTILVLSV 360

Db 301 LLAGVRVAPARSKRWDPKKGTSLSKLSFSTHTLLGQFFQGWGTWVASMPLTILVLSV 360

QY 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFGPPFRTNQVILTAPNRSY 420

Db 361 IPVVALAAGLVFTELTTPVELWSAPNSQARSEKAFHDOHFGPPFRTNQVILTAPNRSY 420

QY 421 RYDLSLLGPKNPSGLTLDLLELLELQBRHLQVWSPQARNISLQDICYAPLNPDNT 480

Db 421 RYDLSLLGPKNPSGLTLDLLELLELQBRHLQVWSPQARNISLQDICYAPLNPDNT 480

QY 481 SLYDCINSLQYFQNNRTHLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKQGTALAL 540

Db 481 SLYDCINSLQYFQNNRTHLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKQGTALAL 540

QY 541 SCWADYGAPVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPPLAQKLWEAEFLSEM 600

Db 541 SCWADYGAPVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPPLAQKLWEAEFLSEM 600

QY 601 RAFORRNAGMFQVTTFAERSLEDEINFTTAEDLPIFATSYIVIFLYTSLALGSYSSRSRV 660

Db 601 RAFORRNAGMFQVTTFAERSLEDEINFTTAEDLPIFATSYIVIFLYTSLALGSYSSRSRV 660

QY 661 MVDKATLGLGGVAVVLGAVMAAGFFSYLGIRSSLVILQVVPFLVLSVGNADNIFIFVLE 720

Db 661 MVDKATLGLGGVAVVLGAVMAAGFFSYLGIRSSLVILQVVPFLVLSVGNADNIFIFVLE 720

QY 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSRAICFFLIGALTPMPAVRTFALTSLGLAV 780

Db 721 YQRLPRRPGPREVHIGRALGRVAPSMMLCSLSRAICFFLIGALTPMPAVRTFALTSLGLAV 780

QY 781 ILDELLQMSAFVALLSLDSKROEASRLDVCCKVQPOELPRGCGEGLLIGFFQKAYAPFL 840

Db 781 ILDELLQMSAFVALLSLDSKROEASRLDVCCKVQPOELPRGCGEGLLIGFFQKAYAPFL 840

QY 841 LHWITRGVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900

Db 841 LHWITRGVLLLFALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900

QY 901 YFVTTLYNFSSEAGMNAICSSAGCNPFTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960

Db 901 YFVTTLYNFSSEAGMNAICSSAGCNPFTQKIQYATEPPEQSYLAIPASSWVDDFIDW 960

QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMISITMGSVRPSVEQFHKYLPMFLNDRP 1020

Db 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKCMISITMGSVRPSVEQFHKYLPMFLNDRP 1020

QY 1021 NIKCPKGLAAYSTSVNLTSDGQVLASRFMAYHKPKNSQDYTEALRAARELANITADL 1080

Db 1021 NIKCPKGLAAYSTSVNLTSDGQVLASRFMAYHKPKNSQDYTEALRAARELANITADL 1080

QY 1081 RKVPGTDPAPVEFPYTTITNVFYQYLTILPEGLFMLSCLCLVPTFAVSCLLGLDLSGLL 1140

Db 1081 RKVPGTDPAPVEFPYTTITNVFYQYLTILPEGLFMLSCLCLVPTFAVSCLLGLDLSGLL 1140

QY 1141 NLLSIVMLVDTVGFMAWDISYNAVSLINLVSAGVMSVVFVSHITRSPAI STKPTWLER 1200

Db 1141 NLLSIVMLVDTVGFMAWDISYNAVSLINLVSAGVMSVVFVSHITRSPAI STKPTWLER 1200

QY 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQIQLIFFRLNLLITLLGLLHGLVFLPV 1260

Db 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQIQLIFFRLNLLITLLGLLHGLVFLPV 1260

QY 1261 ILSVVGDPVNPALALEOKRAEBAVAAVMVASCPNHPRSVSTADNIYNNHSPGSIKAGA 1320

Db 1261 ILSVVGDPVNPALALEOKRAEBAVAAVMVASCPNHPRSVSTADNIYNNHSPGSIKAGA 1320

QY 1321 ISNFPNNGRQF 1332

Db 1321 ISNFPNNGRQF 1332

RESULT 7

US-10-736-769-4

; Sequence 4, Application US/10736769

; GENERAL INFORMATION:

; APPLICANT: Altmann, Scott W

; APPLICANT: Wang, Luquan

; APPLICANT: Graziano, Michael

; APPLICANT: Murgolo, Nick

; TITLE OF INVENTION: NPC111 (NPC3) AND METHODS OF USE THEREOF

; FILE REFERENCE: JB01603-K3-US

; CURRENT APPLICATION NUMBER: US/10/736,769

; CURRENT FILING DATE: 2003-12-16

; PRIOR APPLICATION NUMBER: 60/397,442

; PRIOR FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 10/621,758

; PRIOR FILING DATE: 2003-07-17

; PRIOR APPLICATION NUMBER: 10/646,301

; PRIOR FILING DATE: 2003-08-22

; PRIOR APPLICATION NUMBER: 10/663,208

; PRIOR FILING DATE: 2003-09-16

; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-736-769-4

Query Match 100.0%; Score 6909; DB 37; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEAGLRGWLWALLRLAQSEPTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN 60  
Db 1 MAEAGLRGWLWALLRLAQSEPTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN 60

Qy 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLASISITKALLTRCPACSDNF 120

Qy 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRVPAA 180  
Db 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRVPAA 180

Qy 181 ATLAGVTMCGYVSGALCNAQRWLNFGQDTGNGLAPLDITFHLLEPGQAVSGIQPLNEGV 240  
Db 181 ATLAGVTMCGYVSGALCNAQRWLNFGQDTGNGLAPLDITFHLLEPGQAVSGIQPLNEGV 240

Qy 241 ARCNEOGDDVATCSCODCAASCPAIARPOLDSTFVLGOMPGSLVLIILCSVFVAVTI 300  
Db 241 ARCNEOGDDVATCSCODCAASCPAIARPOLDSTFVLGOMPGSLVLIILCSVFVAVTI 300

Qy 301 LLVGFVRVAPARDKSMVDPKKGTSLSDKLSFSTHLLAGQFFQGGTWTWASWPLTILVLSV 360  
Db 301 LLVGFVRVAPARDKSMVDPKKGTSLSDKLSFSTHLLAGQFFQGGTWTWASWPLTILVLSV 360

Qy 361 IPVVALAAGLVFTELTDPVELSAPNSQARSEKAFHDQHPGFPFRTNQVILTAPRSSY 420  
Db 361 IPVVALAAGLVFTELTDPVELSAPNSQARSEKAFHDQHPGFPFRTNQVILTAPRSSY 420

Qy 421 RYDSLILGPKNFSGILDLLLELLEQLERHLQVWSPQARNISLQDICYAPLNPDNT 480  
Db 421 RYDSLILGPKNFSGILDLLLELLEQLERHLQVWSPQARNISLQDICYAPLNPDNT 480

Qy 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVMDKDFLYCANAPLTFKDGATLAL 540  
Db 481 SLYDCCINSILQYFQNNRTLLLTANQTLMGQTSQVMDKDFLYCANAPLTFKDGATLAL 540

Qy 541 SCHADYGAPVFPPLAIGGYKGDYSEBALIMTFSLNYPAGDPRLAQAKLWEEAFLEEM 600  
Db 541 SCHADYGAPVFPPLAIGGYKGDYSEBALIMTFSLNYPAGDPRLAQAKLWEEAFLEEM 600

Qy 601 RAFQRMAGFQVFTFAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSSYSWSRV 660  
Db 601 RAFQRMAGFQVFTFAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSSYSWSRV 660

Qy 661 MVDKATLGLGGVAVVLGAWAAGFFSYLGRSSVLVQVFPFLVLSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVVLGAWAAGFFSYLGRSSVLVQVFPFLVLSVGADNIFIFVLE 720

Qy 721 YQRLPRRPGPRVHIGRALGRVAPSWMLCSLSEACFFLIGALTTPMPAVRTFALTSGLAV 780  
Db 721 YQRLPRRPGPRVHIGRALGRVAPSWMLCSLSEACFFLIGALTTPMPAVRTFALTSGLAV 780

Qy 781 ILDFLOMSAFVALLSLSKQSEASLDVCCCKVQBELPPPGOGEGILLIGFFOKAYAPFL 840  
Db 781 ILDFLOMSAFVALLSLSKQSEASLDVCCCKVQBELPPPGOGEGILLIGFFOKAYAPFL 840

Qy 841 LHWITRGVLLLFALFGVSLYSNMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSNMCHI SVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900

Qy 901 YFVTTLTGYNFSSBAGNNAICSSAGCANNFSFTQKIYATEPPEQSYLAIPASSWVDDFDIW 960

Db 901 YFVTTLTGYNFSSBAGNNAICSSAGCANNFSFTQKIYATEPPEQSYLAIPASSWVDDFDIW 960

Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCUKNCSITMGSVRPSVEQPHKYLPMFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCUKNCSITMGSVRPSVEQPHKYLPMFLNDRP 1020

Qy 1021 NIKCPKGGLAAYSTSVNLTSDGQVLAASRFMAYHKPKLNQSDYTEALAAARELANITADL 1080  
Db 1021 NIKCPKGGLAAYSTSVNLTSDGQVLAASRFMAYHKPKLNQSDYTEALAAARELANITADL 1080

Qy 1081 RKVPGETDPAPEVEPYTITNVFYEQYLTIPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGETDPAPEVEPYTITNVFYEQYLTIPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140

Qy 1141 NLLSIVMILVDVTGFMALWDISYNAVSLINLVSAGVMSVEFVSHITRSPAIKPTWLER 1200  
Db 1141 NLLSIVMILVDVTGFMALWDISYNAVSLINLVSAGVMSVEFVSHITRSPAIKPTWLER 1200

Qy 1201 AKEATISMGSAVPAVAMTNLPGILVLGLAKAQLIQIFFPRLNLLITLLGLLHGLVLPV 1260  
Db 1201 AKEATISMGSAVPAVAMTNLPGILVLGLAKAQLIQIFFPRLNLLITLLGLLHGLVLPV 1260

Qy 1261 ILSYVGPDPNPALALQKRAEBAVAAVMVASCNPHEFSRVSTADNIYVNHSEFSGIKAGA 1320  
Db 1261 ILSYVGPDPNPALALQKRAEBAVAAVMVASCNPHEFSRVSTADNIYVNHSEFSGIKAGA 1320

Qy 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

RESULT 8  
US-10-750-386-4  
; Sequence 4: Application US/10750386  
; GENERAL INFORMATION:  
; APPLICANT: Garcia-Calvo, Margarita  
; APPLICANT: Chapman, Kevin  
; APPLICANT: Goulet, Mark  
; APPLICANT: Ujjainwalla, Peroze  
; APPLICANT: Altmann, Scott W  
; APPLICANT: Davis, Chip  
; APPLICANT: Bull, Herb  
; APPLICANT: Thornberry, Nancy A  
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF IDENTIFYING AGONISTS AND ANTAGONISTS  
; FILE REFERENCE: A36104 074669.0103  
; CURRENT APPLICATION NUMBER: US/10750,386  
; CURRENT FILING DATE: 2003-12-31  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-750-386-4

Query Match 100.0%; Score 6909; DB 37; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEAGLRGWLWALLRLAQSEPTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN 60  
Db 1 MAEAGLRGWLWALLRLAQSEPTTHQGYCAFYDECGKNPELSGLMTLSNVCSLSN 60

Qy 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLASISITKALLTRCPACSDNF 120

Qy 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRVPAA 180  
Db 121 VNLHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRVPAA 180

Qy 181 ATLAGVTMCGYVSGALCNAQRWLNFGQDTGNGLAPLDITFHLLEPGQAVSGIQPLNEGV 240

Db 181 ATLAAGTMCYGGALCNAQRWLNFGDGTGNGLAPLDITFHLBPQAVGSGIQLNEGV 240  
Qy 241 ARCNEQDDVATCSCDCAACPAIARPAQALDSTFYLGQMPGSLVLIILCSFVAVVTI 300  
Db 241 ARCNEQDDVATCSCDCAACPAIARPAQALDSTFYLGQMPGSLVLIILCSFVAVVTI 300  
Qy 301 LLVGRVAPARDKSMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASHPLTILVLSV 360  
Db 301 LLVGRVAPARDKSMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASHPLTILVLSV 360  
Qy 361 IPVVALAAGLVFTLTDPVLMWAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSSY 420  
Db 361 IPVVALAAGLVFTLTDPVLMWAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSSY 420  
Qy 421 RYDSLILGPKNFGSGLDLDLLELELERLRLHQLVMSPEAQRNLSIODICYAPLNPDT 480  
Db 421 RYDSLILGPKNFGSGLDLDLLELELERLRLHQLVMSPEAQRNLSIODICYAPLNPDT 480  
Qy 481 SLYDCCINSLLQYFONNETLALLTANOTLMGOTSQVDWKDHFLYCANAPLTFKDGTAAL 540  
Db 481 SLYDCCINSLLQYFONNETLALLTANOTLMGOTSQVDWKDHFLYCANAPLTFKDGTAAL 540  
Qy 541 SCMDYGAQVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPAGDPRLAQAKLWEAFLEEM 600  
Db 541 SCMDYGAQVPFPFLAIGYKKGKDYSEAEALIMTFSLNYPAGDPRLAQAKLWEAFLEEM 600  
Qy 601 RAPQRMAGMFOVTPTAERSLEDEINRTTAEDLPIFATSYIVIFLYISALGYSVSSWRV 660  
Db 601 RAPQRMAGMFOVTPTAERSLEDEINRTTAEDLPIFATSYIVIFLYISALGYSVSSWRV 660  
Qy 661 MVDSKATGLGGAUVVLGAVMAAGFFSVGLTRSSLVILQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MVDSKATGLGGAUVVLGAVMAAGFFSVGLTRSSLVILQVVPFLVLSVGADNIFIFVLE 720  
Qy 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSISEAICFFLGALTTPMPAVRTFALTSLGLAV 780  
Db 721 YQRLPRRPGEPREVIHGRALGRVAPSMLLCSISEAICFFLGALTTPMPAVRTFALTSLGLAV 780  
Qy 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCCKVQPELPPQGGEGLLLGFFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKQAEASRLDVCCCKVQPELPPQGGEGLLLGFFQKAYAPFL 840  
Qy 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSVLLDYFLFLNRYEPVGA 900  
Db 841 LHWITRGVLLLFALFGVSLYSMCHI SVGLDQELALPKDSVLLDYFLFLNRYEPVGA 900  
Qy 901 YFVTTILGYNFSSAGMNAICSSAGCNFSTQKI QYATEFFPEQSYLAIPASSWVDDFD 960  
Db 901 YFVTTILGYNFSSAGMNAICSSAGCNFSTQKI QYATEFFPEQSYLAIPASSWVDDFD 960  
Qy 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCNMSITWGSVRPSVEQPHKYL PWFNLDRP 1020  
Db 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCNMSITWGSVRPSVEQPHKYL PWFNLDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSQVLAASFMAVHKPLKNSQDYTEALRAARELAANITADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLTSQVLAASFMAVHKPLKNSQDYTEALRAARELAANITADL 1080  
Qy 1081 RKVPGTDPAFEPFVTITNPFYEQYLTILPEGLFMLSCLVPTPAVSCLLGLDLSRGLL 1140  
Db 1081 RKVPGTDPAFEPFVTITNPFYEQYLTILPEGLFMLSCLVPTPAVSCLLGLDLSRGLL 1140  
Qy 1141 NLLSVIMLVDTVGFMAWLDISYNAVSLINLVSAVGMSEFVSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSVIMLVDTVGFMAWLDISYNAVSLINLVSAVGMSEFVSHITRSFAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAGVAMTNLPGIILVGLAKAQLQIIPFFRLNLIITLLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAGVAMTNLPGIILVGLAKAQLQIIPFFRLNLIITLLGLLHGLVFLPV 1260  
Qy 1261 ILSYVGPDVNPALAEQKRAEEAVAAVMVASCPNHPSRVSTADNIYVNHSPFGSIKGAGA 1320

Db 1261 ILSYVGPDVNPALAEQKRAEEAVAAVMVASCPNHPSRVSTADNIYVNHSPFGSIKGAGA 1320  
Qy 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332  
RESULT 9  
PCT-US01-04098A-1831  
; Sequence 1831, Application PC/TUS0104098A  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-029  
; CURRENT APPLICATION NUMBER: PCT/US01/04098A  
; CURRENT FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: Not Yet Assigned  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 09/728,422  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 09/693,325  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/663,561  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/654,936  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/620,325  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/598,075  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 3960  
; SOFTWARE: Custom  
; SEQ ID NO 1831  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US01-04098A-1831  
Query Match 99.8%; Score 6896; DB 1; Length 1332;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 MAEAGLRGWLWALLRLAQSEPYTTHQPGYCAFYDECGKNPELSGLMTLNVCLSN 60  
Db 1 MAEAGLRGWLWALLRLAQSEPYTTHQPGYCAFYDECGKNPELSGLMTLNVCLSN 60  
Qy 61 TPARKITGDHLLILLOKICPLRYTGPNTQACCSAKQLVSEASLSITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPLRYTGPNTQACCSAKQLVSEASLSITKALLTRCPACSDNF 120  
Qy 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFYQHSFAEQSYDCSRVRVPA 180  
Db 121 VNLHCHNTCSNQSLFINVTRVAQLGAGQLPAVVAEYAFYQHSFAEQSYDCSRVRVPA 180  
Qy 181 ATLAAGTMCYGGALCNAQRWLNFGDGTGNGLAPLDITFHLBPQAVGSGIQLNEGV 240  
Db 181 ATLAAGTMCYGGALCNAQRWLNFGDGTGNGLAPLDITFHLBPQAVGSGIQLNEGV 240  
Qy 241 ARCNEQDDVATCSCDCAACPAIARPAQALDSTFYLGQMPGSLVLIILCSFVAVVTI 300  
Db 241 ARCNEQDDVATCSCDCAACPAIARPAQALDSTFYLGQMPGSLVLIILCSFVAVVTI 300  
Qy 301 LLVGRVAPARDKSMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASHPLTILVLSV 360  
Db 301 LLVGRVAPARDKSMVDPKKGTSLSDKLSFSTHTLLGQFQCGWGTWVASHPLTILVLSV 360  
Qy 361 IPVVALAAGLVFTLTDPVLMWAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSSY 420  
Db 361 IPVVALAAGLVFTLTDPVLMWAPNSQARSEKAFHDQHPGPFRTNQVILTAPNRSSY 420

```
QY 421 RYDSLLGPKNFSGIILDLDDLLLELQERLRHLQVWSPQAQRNLSLQDICYAPLNPNT 480
DB 421 RYDSLLGPKNFSGIILDLDDLLLELQERLRHLQVWSPQAQRNLSLQDICYAPLNPNT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTAAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTAAL 540
QY 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
DB 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
QY 601 RAPQRMAGMFQVTFPAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660
DB 601 RAPQRMAGMFQVTFPAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660
QY 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLVILQVVPFLVLSVGADNIFIVLE 720
DB 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLVILQVVPFLVLSVGADNIFIVLE 720
QY 721 YQRLPRRPGEPREHVHIGRALGRVAPSMMLCSLSEALCFFLGALTMPAVRTFALTSGLA 780
DB 721 YQRLPRRPGEPREHVHIGRALGRVAPSMMLCSLSEALCFFLGALTMPAVRTFALTSGLA 780
QY 781 ILDFLLQMSAFVALLSLDSKQDEASRLDVCCCVKQPQLPPPGQEGGLLGGFFQKAYAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSKQDEASRLDVCCCVKQPQLPPPGQEGGLLGGFFQKAYAPFL 840
QY 841 LHWITRGVLLPLFALFGVSLYSMCHISVGLDQELALPKDSYLLDFLFLNRYFEVGAPV 900
DB 841 LHWITRGVLLPLFALFGVSLYSMCHISVGLDQELALPKDSYLLDFLFLNRYFEVGAPV 900
QY 901 YFVTTLGYNFSSAGNNAICSSAGCNCFSTQKIQVATEFPPEQSYLAIIPASSWDDFIDW 960
DB 901 YFVTTLGYNFSSAGNNAICSSAGCNCFSTQKIQVATEFPPEQSYLAIIPASSWDDFIDW 960
QY 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSITMGSVRPSVEQPHKYLFPWFLNDRP 1020
DB 961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCNKMSITMGSVRPSVEQPHKYLFPWFLNDRP 1020
QY 1021 NIKCKPGGLAAYSTVNLSDQVLAASRPMAYHKPKNSQDTEALRAARELAANITADL 1080
DB 1021 NIKCKPGGLAAYSTVNLSDQVLAASRPMAYHKPKNSQDTEALRAARELAANITADL 1080
QY 1081 RKVPGTDPAFEFPPYITINVFYEOYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140
DB 1081 RKVPGTDPAFEFPPYITINVFYEOYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL 1140
QY 1141 NLLSIVMLIVDTVGFMAWLDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200
DB 1141 NLLSIVMLIVDTVGFMAWLDISYNAVSLINLVSAGMSVEFVSHITRSFAISTKPTWLER 1200
QY 1201 AKEATISMGSAVAGVAMTNLPGIILVLGLAKAQLIQIFPFRNLITLGLLHGLVFLPV 1260
DB 1201 AKEATISMGSAVAGVAMTNLPGIILVLGLAKAQLIQIFPFRNLITLGLLHGLVFLPV 1260
QY 1261 ILSYVGPDVNPALAEOKRAEBAVAAMVASCPNHPRSVSTADNIVVNHSPGSIKGAGA 1320
DB 1261 ILSYVGPDVNPALAEOKRAEBAVAAMVASCPNHPRSVSTADNIVVNHSPGSIKGAGA 1320
QY 1321 ISNFLPNNGRQF 1332
DB 1321 ISNFLPNNGRQF 1332
```

RESULT 10

US-10-239-316-9  
; Sequence 9, Application US/10239316  
; GENERAL INFORMATION:  
; APPLICANT: TANIYAMA, Yoshio  
; APPLICANT: KITA, Shunbun  
; APPLICANT: SATOMI, Tomoko Komiyama

; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof  
; FILE REFERENCE: 2703USOP  
; CURRENT APPLICATION NUMBER: US/10/239,316  
; CURRENT FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: PCT/JP01/02279  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: JP2000-088595  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 59  
; SEQ ID NO 9  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Human  
; US-10-239-316-9

Query Match 99.8%; Score 6896; DB 32; Length 1332;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MAEAGLRGWLWALLRLLAQSEPYTTIHQGYCAFYDECGKPELSGSLMTLSNVCLSN 60
DB 1 MAEAGLRGWLWALLRLLAQSEPYTTIHQGYCAFYDECGKPELSGSLMTLSNVCLSN 60
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLSEASITKALLTRCPACSDNF 120
DB 61 TPARKITGDHLLILLOKICPRLYTGPNTQACCSAKQLVSLSEASITKALLTRCPACSDNF 120
QY 121 VNLHCHNTCSNQSLFINVTRVAQLGAGOLPAVAYEAFYQHSFAESYDSCSRVRPAA 180
DB 121 VNLHCHNTCSNQSLFINVTRVAQLGAGOLPAVAYEAFYQHSFAESYDSCSRVRPAA 180
QY 181 ATLAVTGTCGVYGSALCNAQRWLNFGQDTGNGLAPLDITFHLLPEQOAVSGIQPLNEGV 240
DB 181 ATLAVTGTCGVYGSALCNAQRWLNFGQDTGNGLAPLDITFHLLPEQOAVSGIQPLNEGV 240
QY 241 ARCNSQGGDDVATCSCQDCAASCPAIPALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
DB 241 ARCNSQGGDDVATCSCQDCAASCPAIPALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
QY 301 LLVGRFVAPARDKSKWDPKKGTSISDKLSFSHTHTLLGQFQGWGTWASWPLTILVLSV 360
DB 301 LLVGRFVAPARDKSKWDPKKGTSISDKLSFSHTHTLLGQFQGWGTWASWPLTILVLSV 360
QY 361 IPWVALAAGLVFETLTTPVELWSPNSQARSEKAFHDQHFQFFFRNTQVILITAPNRSSY 420
DB 361 IPWVALAAGLVFETLTTPVELWSPNSQARSEKAFHDQHFQFFFRNTQVILITAPNRSSY 420
QY 421 RYDSLLGPKNFSGIILDLDDLLLELQERLRHLQVWSPQAQRNLSLQDICYAPLNPNT 480
DB 421 RYDSLLGPKNFSGIILDLDDLLLELQERLRHLQVWSPQAQRNLSLQDICYAPLNPNT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTAAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDKDHFLYCANAPLTFKDGTAAL 540
QY 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
DB 541 SCWADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
QY 601 RAPQRMAGMFQVTFPAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660
DB 601 RAPQRMAGMFQVTFPAERSLEDEINRTTAEDLPIFATSYIVIFLYISLALGSYSWSRV 660
QY 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLVILQVVPFLVLSVGADNIFIVLE 720
DB 661 MYDSKATLGLGGVAVVLGAVMAAMGFFSYLGRSSVLVILQVVPFLVLSVGADNIFIVLE 720
QY 721 YQRLPRRPGEPREHVHIGRALGRVAPSMMLCSLSEALCFFLGALTMPAVRTFALTSGLA 780
DB 721 YQRLPRRPGEPREHVHIGRALGRVAPSMMLCSLSEALCFFLGALTMPAVRTFALTSGLA 780
QY 781 ILDFLLQMSAFVALLSLDSKQDEASRLDVCCCVKQPQLPPPGQEGGLLGGFFQKAYAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSKQDEASRLDVCCCVKQPQLPPPGQEGGLLGGFFQKAYAPFL 840
```



Db 781 ILDFLOMSAFVALLSLDSKRQASRLDVCCVKPQELPPPGQEGILLGLFFQKAYAPFL 840  
Qy 841 LHWITRGVLLFLALFGVSLYSNMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGAPV 900  
Db 841 LHWITRGVLLFLALFGVSLYSNMCHISVGLDQELALPKDSYLLDYFLFLNRYFVGAPV 900  
Qy 901 YFVTTLGYNFSSEAGNNAICSSAGCNFSTOKIOVATEPEOSYLAIPASSWVDDFDW 960  
Db 901 YFVTTLGYNFSSEAGNNAICSSAGCNFSTOKIOVATEPEOSYLAIPASSWVDDFDW 960  
Qy 961 LTPSSCRLYISGNKDKFCPSTVNSLNCIKNCMSITMGSVRPSVEQFKYLPWFNDRP 1020  
Db 961 LTPSSCRLYISGNKDKFCPSTVNSLNCIKNCMSITMGSVRPSVEQFKYLPWFNDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFYHAYHKLKNSQDYTEALRAARELANITADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFYHAYHKLKNSQDYTEALRAARELANITADL 1080  
Qy 1081 RKVPGTDPAFVFPYTTITNVFYEQYLTILPEGLFMLSCLVPTPAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFVFPYTTITNVFYEQYLTILPEGLFMLSCLVPTPAVSCLLGLDLRSGLL 1140  
Qy 1141 NLLSIVMLVDVTFGMALWDISYNAVSLINLVSAGMSVDFVSHITRSFAISTKPTWLER 1200  
Db 1141 NLLSIVMLVDVTFGMALWDISYNAVSLINLVSAGMSVDFVSHITRSFAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAVAGVAMTNLPGILVLGLAKAQIQTFFRNLNLLITLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAVAGVAMTNLPGILVLGLAKAQIQTFFRNLNLLITLGLLHGLVFLPV 1260  
Qy 1261 ILSYVGPDPNPALEOKRAEBAEAAVAVVASCNHPHSRVSTADNIIYNHSPFEGSIKAGA 1320  
Db 1261 ILSYVGPDPNPALEOKRAEBAEAAVAVVASCNHPHSRVSTADNIIYNHSPFEGSIKAGA 1320  
Qy 1321 ISNFLPNNGRQF 1332  
Db 1321 ISNFLPNNGRQF 1332

RESULT 11

US-10-258-899A-1831  
; Sequence 1831, Application US/10258899A  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Drmanac, Radoje T.  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhou, Ping  
; APPLICANT: Xu, Chongjun  
; APPLICANT: Cao, Yicheng  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Wang, Zhi Wei  
; APPLICANT: Xue, Aidong  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Goodrich, Ryle  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 787CIP2-2G/US  
; CURRENT APPLICATION NUMBER: US/10/258,899A  
; CURRENT FILING DATE: 2003-11-06  
; PRIOR APPLICATION NUMBER: PCT/US01/04098  
; PRIOR FILING DATE: 2001-02-05  
; PRIOR APPLICATION NUMBER: 09/774,434  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 09/728,422  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 09/693,325

; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/663,561  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/654,936  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/620,325  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/598,075  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 3960  
; SOFTWARE: Custom  
; SEQ ID NO 1831  
; TYPE: PRT  
; LENGTH: 1332  
; ORGANISM: Homo sapiens  
; US-10-258-899A-1831  
  
Query Match 99.8%; Score 6896; DB 32; Length 1332;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 MAEAGLRGWLWALLRLAQSEPYTHQPGYCAFYDECGKNPELSGSLMTLSNVCSLSN 60  
Db 1 MAEAGLRGWLWALLRLAQSEPYTHQPGYCAFYDECGKNPELSGSLMTLSNVCSLSN 60  
Qy 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF 120  
Qy 121 VNLHCHNTCSPNOSLFNTRVAQLGAGOLPAVVAEAFYQHSFAQSQSDSCSRVRPAA 180  
Db 121 VNLHCHNTCSPNOSLFNTRVAQLGAGOLPAVVAEAFYQHSFAQSQSDSCSRVRPAA 180  
Qy 181 ATLAAGTMCVYGSALCNARWLNFQDGTGNGLAPLDITPHLLPEGAQVSGIQLNEGV 240  
Db 181 ATLAAGTMCVYGSALCNARWLNFQDGTGNGLAPLDITPHLLPEGAQVSGIQLNEGV 240  
Qy 241 ARNESQGDVATCSQDCAASCPATARPQALDSTFVGMPGSLVLIITLCSVFAVVTI 300  
Db 241 ARNESQGDVATCSQDCAASCPATARPQALDSTFVGMPGSLVLIITLCSVFAVVTI 300  
Qy 301 LLVGRVAPARDKSKMVDPKKGTSLSDKLSFSFTHLLGQFFQCGMTWVASWPLTILVLSV 360  
Db 301 LLVGRVAPARDKSKMVDPKKGTSLSDKLSFSFTHLLGQFFQCGMTWVASWPLTILVLSV 360  
Qy 361 IPVVALAAGLVFTLTTPVELMSAPNSQARSEKAFHQHFGFPFRTNQVILTAPNRSY 420  
Db 361 IPVVALAAGLVFTLTTPVELMSAPNSQARSEKAFHQHFGFPFRTNQVILTAPNRSY 420  
Qy 421 RYDSSLILGPKNFGSILDLILLLELELOERLHLQVWSPQAQRNISIQDICYAPLNPDNT 480  
Db 421 RYDSSLILGPKNFGSILDLILLLELELOERLHLQVWSPQAQRNISIQDICYAPLNPDNT 480  
Qy 481 SLXDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGTAAL 540  
Db 481 SLXDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGTAAL 540  
Qy 541 SCMDYGAAPVFPFLAIGYKGYKDYSEAEALIMTSLNYPAGDPRLAQAKLWEAFLEEM 600  
Db 541 SCMDYGAAPVFPFLAIGYKGYKDYSEAEALIMTSLNYPAGDPRLAQAKLWEAFLEEM 600  
Qy 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSSSSRSR 660  
Db 601 RAFQRMAGMFQVTFMAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSSSSRSR 660  
Qy 661 MVDSKATLGLGVAVVLGAVNAAMGFFSYLGRSSVLQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MVDSKATLGLGVAVVLGAVNAAMGFFSYLGRSSVLQVVPFLVLSVGADNIFIFVLE 720





Db 901 YFVTLLGYNFSSEAGMNAICSSAGCNFFSTQKIQYATEPEQSYLAIPASSWDDFDW 960  
Qy 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMCSITMGSVRPSVEQFHXYLPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMCSITMGSVRPSVEQFHXYLPWFLNDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFVAYHKKPLKNSQDYTEALRAARELANITADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFVAYHKKPLKNSQDYTEALRAARELANITADL 1080  
Qy 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSLCLVPTPAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSLCLVPTPAVSCLLGLDLRSGLL 1140  
Qy 1141 NLSIVMLVDTVGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTWLER 1200  
Db 1141 NLSIVMLVDTVGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTWLER 1200  
Qy 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQILQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
Db 1201 AKEATISMGSAVFAGVAMTNLPGILVLGLAKAQILQIIFFRNLNLLITLLGLLHGLVFLPV 1260  
Qy 1261 ILSYVGPDVNPALAEQKRAEEAAVAVVAVASCPNHPRSVSTADNIYVNHSPFEGSIKAGA 1320  
Db 1261 ILSYVGPDVNPALAEQKRAEEAAVAVVAVASCPNHPRSVSTADNIYVNHSPFEGSIKAGA 1320  
Qy 1321 ISNPLPNNGRQF 1332  
Db 1321 ISNPLPNNGRQF 1332

RESULT 13

US-11-218-141-1831  
; Sequence 1831, Application US/11218141  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y. Tom et al  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-029  
; CURRENT APPLICATION NUMBER: US/11/218,141  
; CURRENT FILING DATE: 2005-08-31  
; PRIOR APPLICATION NUMBER: Not Yet Assigned  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 09/728,422  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: 09/693,325  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/663,561  
; PRIOR FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/654,936  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 09/620,325  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/598,075  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 3960  
; SOFTWARE: Custom  
; SEQ ID NO 1831  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-218-141-1831

Query Match 99.8%; Score 6896; DB 42; Length 1332;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MAEAGLRGWLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60

Db 1 MAEAGLRGWLWALLRLAQSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60  
Qy 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKOLVSEASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLILLOKICPRLYTGPNTOACCSAKOLVSEASISITKALLTRCPACSDNF 120  
Qy 121 VNLHCHTSPNOSLFINTVRVAQLGAGOLPAVVAYEAFYQHSFAEQSYDSCSRVRVPA 180  
Db 121 VNLHCHTSPNOSLFINTVRVAQLGAGOLPAVVAYEAFYQHSFAEQSYDSCSRVRVPA 180  
Qy 181 ATLAVGTMCVYGSALCNAQRWLNFGQDTCNGGLAPLDITPHLLRPGQAVSGGTOPLNEGV 240  
Db 181 ATLAVGTMCVYGSALCNAQRWLNFGQDTCNGGLAPLDITPHLLRPGQAVSGGTOPLNEGV 240  
Qy 241 ARCNESOGDDVATCSQDCAASCPAARPOLADOSTFYLQMPGSLVLIILCSVFAVVTI 300  
Db 241 ARCNESOGDDVATCSQDCAASCPAARPOLADOSTFYLQMPGSLVLIILCSVFAVVTI 300  
Qy 301 LLVGFVRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGMTWVASWPLTILVLSV 360  
Db 301 LLVGFVRVAPARDKSKMVDPKKGTSLSDKLSFSTHTLLGQFFQCGMTWVASWPLTILVLSV 360  
Qy 361 IPVVALAAGLVFTLTTDPVELWSAPNSQARSKAFHDQHFQFFRTNQVILITAPNRSY 420  
Db 361 IPVVALAAGLVFTLTTDPVELWSAPNSQARSKAFHDQHFQFFRTNQVILITAPNRSY 420  
Qy 421 RYDSLILGPKNFSGIILDLLELLELOERLHLQVWSPQAORNIQLDICIYAPLNPDNT 480  
Db 421 RYDSLILGPKNFSGIILDLLELLELOERLHLQVWSPQAORNIQLDICIYAPLNPDNT 480  
Qy 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKDGTAAL 540  
Db 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKDGTAAL 540  
Qy 541 SCWADYGAPVFPFLAIGGYKGDYSEBALIMTFSLNNYPAGDRLAQAOLWEAFLEEM 600  
Db 541 SCWADYGAPVFPFLAIGGYKGDYSEBALIMTFSLNNYPAGDRLAQAOLWEAFLEEM 600  
Qy 601 RAFQRRMAGMFQVTFMAERSLEDEINRTAEDLPIFATSVIVIFLYISLALGSYSSSRV 660  
Db 601 RAFQRRMAGMFQVTFMAERSLEDEINRTAEDLPIFATSVIVIFLYISLALGSYSSSRV 660  
Qy 661 MVDKATLGLGGVAVVLGAVMAAMGFFSYLGIRSSVLQVVPFLVSVGADNIFIFVLE 720  
Db 661 MVDKATLGLGGVAVVLGAVMAAMGFFSYLGIRSSVLQVVPFLVSVGADNIFIFVLE 720  
Qy 721 YQRLPRPGPREVHIGRALGRVAPSMILCSLSEAI CFFILGALTPMPAVRTFALTSGLA 780  
Db 721 YQRLPRPGPREVHIGRALGRVAPSMILCSLSEAI CFFILGALTPMPAVRTFALTSGLA 780  
Qy 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCVKPQELPPPGQEGLLGLGFFQKAYAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCVKPQELPPPGQEGLLGLGFFQKAYAPFL 840  
Qy 841 LHMWITRGVLLLFALFVLSYMSCHISVGLDQBELPKDSYLLDYFLNRYFEVGA 900  
Db 841 LHMWITRGVLLLFALFVLSYMSCHISVGLDQBELPKDSYLLDYFLNRYFEVGA 900  
Qy 901 YFVTTLLGYNFSSEAGMNAICSSAGCNFFSTQKIQYATEPEQSYLAIPASSWDDFDW 960  
Db 901 YFVTTLLGYNFSSEAGMNAICSSAGCNFFSTQKIQYATEPEQSYLAIPASSWDDFDW 960  
Qy 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMCSITMGSVRPSVEQFHXYLPWFLNDRP 1020  
Db 961 LTPSSCCRLYISGNKDKFCPSTVNSLNCNKMCSITMGSVRPSVEQFHXYLPWFLNDRP 1020  
Qy 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFVAYHKKPLKNSQDYTEALRAARELANITADL 1080  
Db 1021 NIKCPKGLAAYSTSVNLTSDQVLAASFVAYHKKPLKNSQDYTEALRAARELANITADL 1080  
Qy 1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSLCLVPTPAVSCLLGLDLRSGLL 1140

Db 1081 RKVPGTDPAFVPPYITINVFYEQYLITLPEGLFMLSCLVLPFAVSCLLGLDLRSGLL 1140  
QY 1141 NLLSIVMILVDTVGFMAWLDISYNAVSLINLVSAGVMSVEFVSHITRSPAISTKPTWLER 1200  
Db 1141 NLLSIVMILVDTVGFMAWLGISYNAVSLINLVSAGVMSVEFVSHITRSPAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVPAVAVMTNLPGLILVLGLAKAQLIQIFFRLNLLITLLGLLHGLVPLPV 1260  
Db 1201 AKEATISMGSAVPAVAVMTNLPGLILVLGLAKAQLIQIFFRLNLLITLLGLLHGLVPLPV 1260  
QY 1261 ILSYVGPDPNPALALEOKRAEEAVAAVMVASCNPNHPSRVSTADNIYVNHSPFSGSIKAGA 1320  
Db 1261 ILSYVGPDPNPALALEOKRAEEAVAAVMVASCNPNHPSRVSTADNIYVNHSPFSGSIKAGA 1320  
QY 1321 ISNPLPNNGRQF 1332  
Db 1321 ISNPLPNNGRQF 1332

RESULT 14  
US-11-242-459-9  
; Sequence 9, Application US/11242459  
; GENERAL INFORMATION:  
; APPLICANT: TANIYAMA, Yoshio  
; APPLICANT: KITA, Shunbun  
; APPLICANT: SATOMI, Tomoko Komiyama  
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof  
; FILE REFERENCE: 2703USOP  
; CURRENT APPLICATION NUMBER: US/11/242,459  
; CURRENT FILING DATE: 2005-10-03  
; PRIOR APPLICATION NUMBER: US/10/239,316  
; PRIOR FILING DATE: 2002-09-19  
; PRIOR APPLICATION NUMBER: PCT/JP01/02279  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: JP2000-088595  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 59  
; SEQ ID NO 9  
; LENGTH: 1332  
; TYPE: PRT  
; ORGANISM: Human  
US-11-242-459-9

Query Match 99.8%; Score 6896; DB 42; Length 1332;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAEAGLRGWLWALLRLAQSEBYTTHQPGYCAFYDECGKNPELSSGLMTLSNVCSLSN 60  
Db 1 MAEAGLRGWLWALLRLAQSEBYTTHQPGYCAFYDECGKNPELSSGLMTLSNVCSLSN 60  
QY 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSLIASISITKALLTRCPACSDNF 120  
Db 61 TPARKITGDHLLILLOKICPRLYTGPNTOACCSAKQLVSLIASISITKALLTRCPACSDNF 120  
QY 121 VNLHCNTCSPNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAPQSYDSCSRVRVPAA 180  
Db 121 VNLHCNTCSPNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAPQSYDSCSRVRVPAA 180  
QY 181 ATLAVTMCVTVGSALCNAQRMNLNFOGDTGNGLAPLDITFHLLEPQAVGSGIQPLNEGV 240  
Db 181 ATLAVTMCVTVGSALCNAQRMNLNFOGDTGNGLAPLDITFHLLEPQAVGSGIQPLNEGV 240  
QY 241 ARCNEQSGDDVATCSCQDCAACPAITARPQALDSTFVLGOMPGLVLIILCSVFAVVTI 300  
Db 241 ARCNEQSGDDVATCSCQDCAACPAITARPQALDSTFVLGOMPGLVLIILCSVFAVVTI 300  
QY 301 LLVGFVRVAPARKSKMVDPKKGTSLSDKLSFSFTHLLGQFFQCGWTWASWPLTILVLSV 360  
Db 301 LLVGFVRVAPARKSKMVDPKKGTSLSDKLSFSFTHLLGQFFQCGWTWASWPLTILVLSV 360  
QY 361 IPVVALAAGLVFTTELTDPVELWSPNSQARSEKAFHQHFGFPFFRTNQVILTAPNRSSY 420

Db 361 IPVVALAAGLVFTTELTDPVELWSPNSQARSEKAFHQHFGFPFFRTNQVILTAPNRSSY 420  
QY 421 RYDSLLGPKNFGSILDLDELLELQERLRLQWMSPEAQRNISLQDICYAPLNPDNT 480  
Db 421 RYDSLLGPKNFGSILDLDELLELQERLRLQWMSPEAQRNISLQDICYAPLNPDNT 480  
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLWGQTSQVDWKDHFHYCANAPLTFKDGATLAL 540  
Db 481 SLYDCCINSLLQYFQNNRTLLLTANQTLWGQTSQVDWKDHFHYCANAPLTFKDGATLAL 540  
QY 541 SCHADYGAPVFPPLATGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAOKLWBEAFLEEM 600  
Db 541 SCHADYGAPVFPPLATGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAOKLWBEAFLEEM 600  
QY 601 RAFQRRMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVFLYISLAGSYSSWSRV 660  
Db 601 RAFQRRMAGMFQVTFTAERSLEDEINRTTAEDLPIFATSYIVFLYISLAGSYSSWSRV 660  
QY 661 MYDSKATLGLGGVAVVVLGAVMAAMGPFYSYLGRSSVLILQVVPFLVLSVGADNIFIFVLE 720  
Db 661 MYDSKATLGLGGVAVVVLGAVMAAMGPFYSYLGRSSVLILQVVPFLVLSVGADNIFIFVLE 720  
QY 721 YQRLPRRPGEPREVHIGRALGRVAPSKMLCSLSEATCFELGALTMPAVRTFALTSGLAV 780  
Db 721 YQRLPRRPGEPREVHIGRALGRVAPSKMLCSLSEATCFELGALTMPAVRTFALTSGLAV 780  
QY 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPPGOGEGLLGLGFFOKAVAPFL 840  
Db 781 ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPPGOGEGLLGLGFFOKAVAPFL 840  
QY 841 LHWITRGVLLLFALFGVSLYSMSCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
Db 841 LHWITRGVLLLFALFGVSLYSMSCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGAPV 900  
QY 901 YFVTTILGYNFSSEAGHNAICSSAGCNPFSTQIKIYATEFPPEQSYLAIPASSWVDDPIDW 960  
Db 901 YFVTTILGYNFSSEAGHNAICSSAGCNPFSTQIKIYATEFPPEQSYLAIPASSWVDDPIDW 960  
QY 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCNKCMSITMGSVRPSVSEQFHKLFWFLNDRP 1020  
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCNKCMSITMGSVRPSVSEQFHKLFWFLNDRP 1020  
QY 1021 NIKCPKGGLAAYSTSVNLTSDGQVLASRFMAHYHKLKNSQDYTEALRAARELANITADL 1080  
Db 1021 NIKCPKGGLAAYSTSVNLTSDGQVLASRFMAHYHKLKNSQDYTEALRAARELANITADL 1080  
QY 1081 RKVPGTDPAFVPPYITINVFYEQYLITLPEGLFMLSCLVLPFAVSCLLGLDLRSGLL 1140  
Db 1081 RKVPGTDPAFVPPYITINVFYEQYLITLPEGLFMLSCLVLPFAVSCLLGLDLRSGLL 1140  
QY 1141 NLLSIVMILVDTVGFMAWLDISYNAVSLINLVSAGVMSVEFVSHITRSPAISTKPTWLER 1200  
Db 1141 NLLSIVMILVDTVGFMAWLDISYNAVSLINLVSAGVMSVEFVSHITRSPAISTKPTWLER 1200  
QY 1201 AKEATISMGSAVPAVAVMTNLPGLILVLGLAKAQLIQIFFRLNLLITLLGLLHGLVPLPV 1260  
Db 1201 AKEATISMGSAVPAVAVMTNLPGLILVLGLAKAQLIQIFFRLNLLITLLGLLHGLVPLPV 1260  
QY 1261 ILSYVGPDPNPALALEOKRAEEAVAAVMVASCNPNHPSRVSTADNIYVNHSPFSGSIKAGA 1320  
Db 1261 ILSYVGPDPNPALALEOKRAEEAVAAVMVASCNPNHPSRVSTADNIYVNHSPFSGSIKAGA 1320  
QY 1321 ISNPLPNNGRQF 1332  
Db 1321 ISNPLPNNGRQF 1332

RESULT 15  
PCT-US01-04098A-1830  
; Sequence 1830, Application PC/TUS0104098A  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides

Mon Apr 10 07:16:15 2006

```
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1830
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1830

Query Match      99.5%; Score 6872.5; DB 1; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

Qy      1 MAEAGLRGWLWALLRLRAQSEPYTHIQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60
Db      1 MAEAGLRGWLWALLRLRAQSEPYTHIQPGYCAFYDECGKNPELSGLMTLSNVCSLSN 60

Qy      61 TPARKITGDHLLLOKICPLRYTGPNTQACSAKOLVLEASLSITKALLTRCPACSNF 120
Db      61 TPARKITGDHLLLOKICPLRYTGPNTQACSAKOLVLEASLSITKALLTRCPACSNF 120

Qy      121 VNLHCHNTCSNQSIFINVTVAQAGOLPAVAYEAFYQHSFAEQSYDSCSRVRVPA 180
Db      121 VNLHCHNTCSNQSIFINVTVAQAGOLPAVAYEAFYQHSFAEQSYDSCSRVRVPA 180

Qy      181 ATLAGVTMGVYGSALCNAQRWLNFGDGTGNGLAPLDITFHLLPEGQAVGSGIQPLNEGV 240
Db      181 ATLAGVTMGVYGSALCNAQRWLNFGDGTGNGLAPLDITFHLLPEGQAVGSGIQPLNEGV 240

Qy      241 ARCNESQGDVATCSCQDCAASPATARPQALDSTFYLGOMPGSLVLIILLCVFAVVTI 300
Db      241 ARCNESQGDVATCSCQDCAASPATARPQALDSTFYLGOMPGSLVLIILLCVFAVVTI 300

Qy      301 LLVGFRVAPARSKMVDPKGTSLSDKLSFSTHLLGQFOGWTWVWASPLTILVLSV 360
Db      301 LLVGFRVAPARSKMVDPKGTSLSDKLSFSTHLLGQFOGWTWVWASPLTILVLSV 360

Qy      361 IPVVALAAGLVFTTELTTDPVELMSAPNSQARSEKAFHDQHFQFPFTNOVILTAPNRSSY 420
Db      361 IPVVALAAGLVFTTELTTDPVELMSAPNSQARSEKAFHDQHFQFPFTNOVILTAPNRSSY 420

Qy      421 RYDSSLGPKNFSGIILDLLELLELLELRLHQLQWSPQARNISLQDICYAPLNPDNT 480
Db      421 RYDSSLGPKNFSGIILDLLELLELLELRLHQLQWSPQARNISLQDICYAPLNPDNT 480

Qy      481 SLVDCCINSLLQVFQNNRTLLLTANOTLMGQTSQVDWKDHFYCANAPLTFKDGFTALAL 540
Db      481 SLVDCCINSLLQVFQNNRTLLLTANOTLMGQTSQVDWKDHFYCANAPLTFKDGFTALAL 540

Qy      541 SCMADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
Db      541 SCMADYGAPVFPFLAIGGYKGDYSEAEALIMTFSLNNYPAGDPRLAQAKLWEEAFLEEM 600
```

```
Qy      601 RAFQRMAGMFOVTFTAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSSWSRV 660
Db      601 RAFQRMAGMFOVTFTAERSLEDEINRTTAEADLPFATSYIVIFLYISLALGSSWSRV 660

Qy      661 MYDSKATLGLGCVAVVLGAVMAAMGFFSYLGRSSILVLOVVPFLVLSVGAONIIFVLE 720
Db      661 MYDSKATLGLGCVAVVLGAVMAAMGFFSYLGRSSILVLOVVPFLVLSVGAONIIFVLE 720

Qy      721 YQRLPRRPGEPREVIHGRALGRVAPSMMLCSLSEAI CFFPLGALTTPMPAVRTFALTSLAV 780
Db      721 YQRLPRRPGEPREVIHGRALGRVAPSMMLCSLSEAI CFFPLGALTTPMPAVRTFALTSLAV 780

Qy      781 ILDFLLQMSAFVALLSLDSKQEAERLDVCCCKVQELPPPGQEGGLLGFQKAYAPFL 840
Db      781 ILDFLLQMSAFVALLSLDSKQEAERLDVCCCKVQELPPPGQEGGLLGFQKAYAPFL 840

Qy      841 LHWITRGVVLLFLALFGVLSYSMCHI SVGLQDELALPKDSYLLDYFLNRYFEVGAPV 900
Db      841 LHWITRGVVLLFLALFGVLSYSMCHI SVGLQDELALPKDSYLLDYFLNRYFEVGAPV 900

Qy      901 YFVTTILGYNFSSEAGMNAICSSAGCNMFSTQKIQYATEFPPEQSYLAI PASSWVDDFIDW 960
Db      901 YFVTTILGYNFSSEAGMNAICSSAGCNMFSTQKIQYATEFPPEQSYLAI PASSWVDDFIDW 960

Qy      961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCCLKMCSITMGSVRPSVQFHKYLPWFNDPR 1020
Db      961 LTPSSCCRLYISGPNKDKFCPSTVNSLNCCLKMCSITMGSVRPSVQFHKYLPWFNDPR 1020

Qy      1021 NIKCPKGGLAAYSTSVNLTSDGOVL-----ASRFMAVH 1053
Db      1021 NIKCPKGGLAAYSTSVNLTSDGOVL-----ASRFMAVH 1053

Qy      1054 KPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPFPYTTITNVFYEQYLTILPEGL 1113
Db      1054 KPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPFPYTTITNVFYEQYLTILPEGL 1113

Qy      1114 FMLSCLCLVPTFAVSCLLGLDLRLSGLNLNLSIVMLVDTVGFMAWDISYNAVSLNLVS 1200
Db      1141 FMLSCLCLVPTFAVSCLLGLDLRLSGLNLNLSIVMLVDTVGFMAWDISYNAVSLNLVS 1200

Qy      1174 AVGMSVEFVSHITRESFAISTKPTWLERAKATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1233
Db      1201 AVGMSVEFVSHITRESFAISTKPTWLERAKATISMGSAVFAGVAMTNLPGILVLGLAKAQ 1260

Qy      1234 LIQIFFFRLNLLITLLGLLHGLVFLPVILSYVGPDPVNPALALEOKRAEEAAVAVNVASCP 1293
Db      1261 LIQIFFFRLNLLITLLGLLHGLVFLPVILSYVGPDPVNPALALEOKRAEEAAVAVNVASCP 1320

Qy      1294 NHPRSVSTADNIYVNHSPFEGSIKAGAI SNFLPNNGRQF 1332
Db      1321 NHPRSVSTADNIYVNHSPFEGSIKAGAI SNFLPNNGRQF 1359
```

Search completed: April 7, 2006, 19:32:35  
Job time : 576 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2006, 19:23:20 ; Search time 77 Seconds  
(without alignments)  
2380.515 Million cell updates/sec

Title: US-10-736-769-4  
Perfect score: 6909  
Sequence: 1 MAEAGLRGWLWALLRLAQ.....GSIKGAGISNFIPLNNGRQF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 584368 seqs, 137612332 residues

Total number of hits satisfying chosen parameters: 584368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New\*

- 1: /SIDS5/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*
- 2: /SIDS5/ptodata/1/paa/US06\_NEW\_COMB.pep.\*
- 3: /SIDS5/ptodata/1/paa/US07\_NEW\_COMB.pep.\*
- 4: /SIDS5/ptodata/1/paa/US08\_NEW\_COMB.pep.\*
- 5: /SIDS5/ptodata/1/paa/US09\_NEW\_COMB.pep.\*
- 6: /SIDS5/ptodata/1/paa/US10\_NEW\_COMB.pep.\*
- 7: /SIDS5/ptodata/1/paa/US11\_NEW\_COMB.pep.\*
- 8: /SIDS5/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	6	US-11-301-094-6
2	6872.5	99.5	1359	6	US-11-270-796-22
3	5407	78.3	1333	6	US-11-270-796-3
4	2402.5	34.8	1278	1	PCT-US06-05584-822
5	2402.5	34.8	1278	6	US-11-191-274A-129
6	2402.5	34.8	1278	6	US-11-191-274A-130
7	2402.5	34.8	1278	7	US-11-385-692-2400
8	2402.5	34.8	1278	7	US-11-385-692-2401
9	1341.5	15.4	1383	6	US-11-301-094-2
10	1046	15.1	1274	6	US-11-301-094-4
11	782.5	11.3	419	6	US-10-953-349-21065
12	588.5	8.5	967	6	US-10-461-673-16739
13	583.5	8.4	1182	6	US-11-332-764-2
14	583.5	8.4	1182	6	US-11-337-244-149
15	482	7.0	891	8	US-60-772-265-1197
16	455	6.6	1358	8	US-60-772-265-241
17	437	6.3	831	7	US-11-360-355-133805
18	349	5.1	783	6	US-10-461-673-16750
19	325	4.7	204	7	US-11-360-355-141351
20	261	3.8	575	7	US-11-360-355-120785
21	252	3.6	505	6	US-10-461-673-16871
22	229.5	3.3	465	7	US-11-360-355-141367
23	225	3.3	542	6	US-11-214-063A-1670
24	222.5	3.2	422	7	US-11-360-355-141365
25	219	3.2	1137	6	US-10-461-673-16889

26	219	3.2	1330	6	US-10-461-673-10541	Sequence 10541, A
27	190	2.8	126	7	US-11-360-355-141352	Sequence 141352
28	152.5	2.2	632	6	US-11-293-697-3548	Sequence 3548, Ap
29	141	2.0	164	7	US-11-360-355-127069	Sequence 127069
30	139.5	2.0	484	7	US-11-360-355-120169	Sequence 120169
31	135.5	2.0	252	7	US-11-360-355-140206	Sequence 140206
32	126	1.8	808	6	US-10-724-972B-7050	Sequence 7050, Ap
33	123.5	1.8	171	7	US-11-360-355-149589	Sequence 149589
34	121.5	1.8	1043	6	US-10-536-606-20	Sequence 20, Appl
35	120.5	1.7	2426	6	US-11-203-806A-11	Sequence 11, Appl
36	118.5	1.7	768	6	US-11-214-063A-2044	Sequence 2044, Ap
37	118.5	1.7	788	6	US-11-214-063A-1692	Sequence 1692, Ap
38	117.5	1.7	619	8	US-60-732-162-1828	Sequence 1828, Ap
39	116.5	1.7	619	6	US-11-312-958-46	Sequence 46, Appl
40	116.5	1.7	619	6	US-10-461-673-12167	Sequence 12167, A
41	114	1.7	488	8	US-60-752-355-45206	Sequence 45206, A
42	114	1.7	985	6	US-11-293-697-2874	Sequence 2874, Ap
43	112.5	1.6	697	6	US-10-703-799B-226	Sequence 226, App
44	111	1.6	201	7	US-11-360-355-152528	Sequence 152528
45	111	1.6	958	8	US-60-781-953-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1

US-11-301-094-6

; Sequence 6, Application US/11301094

; GENERAL INFORMATION:

; APPLICANT: Levitan, Diane J

; APPLICANT: Smith, Marsha

; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS

; FILE REFERENCE: JB06242US01

; CURRENT APPLICATION NUMBER: US/11/301,094

; CURRENT FILING DATE: 2005-12-12

; PRIOR APPLICATION NUMBER: 60/636,390

; PRIOR FILING DATE: 2004-12-15

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 6

; LENGTH: 1332

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-301-094-6

Query Match 100.0%; Score 6909; DB 6; Length 1332;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKNPELSGLMTLSNVCLSN	60
Db	1	MAEAGLRGWLWALLRLAQSEPYTTIHOPGYCAFYDECGKNPELSGLMTLSNVCLSN	60
QY	61	TPARKITGDHLLIQKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF	120
Db	61	TPARKITGDHLLIQKICPRLYTGPNTOACCSAKQLVSLASISITKALLTRCPACSDNF	120
QY	121	VNLHCHTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAPQSDSCSRVRVPA	180
Db	121	VNLHCHTCSNQSLFINVTRVAQLGAGQLPAVAYEAFYQHSFAPQSDSCSRVRVPA	180
QY	181	ATLAVGTCMGVYGSALCNAORWLNFGDGTGNGIAPLIDITFHLLPQVAVGSGIOLNEGV	240
Db	181	ATLAVGTCMGVYGSALCNAORWLNFGDGTGNGIAPLIDITFHLLPQVAVGSGIOLNEGV	240
QY	241	ARCNSQGGDDVATCSQDCAASCPAIPALDFTFYLGQMPGSLVLIILCSVFAVVTI	300
Db	241	ARCNSQGGDDVATCSQDCAASCPAIPALDFTFYLGQMPGSLVLIILCSVFAVVTI	300
QY	301	LLVGRFVAPARDKSNVDPKGTSLSDKLSFSTHTLLGQFPQGWGTWASWPLTILVLSV	360
Db	301	LLVGRFVAPARDKSNVDPKGTSLSDKLSFSTHTLLGQFPQGWGTWASWPLTILVLSV	360



Db	781	ILDLQMSAFVALLSDSKRQASRLDVCCVKPQLPPGQEGGLLGFQKAYAPFL	840
Qy	841	LHWITRGVLLFLALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLNRYFEVCAV	900
Db	841	LHWITRGVLLFLALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLNRYFEVCAV	900
Qy	901	YFVTTLGYNFSSEAGMNAICSSAGCANNFSTQKIQYATEPPEQSILAIPASSWDDPIDW	960
Db	901	YFVTTLGYNFSSEAGMNAICSSAGCANNFSTQKIQYATEPPEQSILAIPASSWDDPIDW	960
Qy	961	LTSSCCRLYISGPNKDKFCPSTVNSLCLNCKMSITMGSVRPSVEQFHKYLWFLNDRP	1020
Db	961	LTSSCCRLYISGPNKDKFCPSTVNSLCLNCKMSITMGSVRPSVEQFHKYLWFLNDRP	1020
Qy	1021	NIKCPKGLAAYSTSVNLTSDGOVL-----ASRFMAYH	1053
Db	1021	NIKCPKGLAAYSTSVNLTSDGOVLDIVALLSPRLEYSGTISAHCNLYLLDLSARFMAYH	1080
Qy	1054	KPLKNSQDYTEALRAARELANITADLRKVPGTDPAPFVPPYITTNVFEQYLTILPEGL	1113
Db	1081	KPLKNSQDYTEALRAARELANITADLRKVPGTDPAPFVPPYITTNVFEQYLTILPEGL	1140
Qy	1114	FMLSCLVPTFAVSCULLGLDLSGLNLSIVMILVDTVGFMAWDISYNAVSLINLVS	1173
Db	1141	FMLSCLVPTFAVSCULLGLDLSGLNLSIVMILVDTVGFMAWDISYNAVSLINLVS	1200
Qy	1174	AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVAGVAMTNLPGLVLGLAKAQ	1233
Db	1201	AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVAGVAMTNLPGLVLGLAKAQ	1260
Qy	1234	LQIOPFRNLNLTLLGLHLGVFLPVILSYVGPDPNPALALBQKRAEBAVAVMVASCP	1293
Db	1261	LQIOPFRNLNLTLLGLHLGVFLPVILSYVGPDPNPALALBQKRAEBAVAVMVASCP	1320
Qy	1294	NHPSRVSTADNIYVNHSPSGIKAGAIISNFLPNNGRQF	1332
Db	1321	NHPSRVSTADNIYVNHSPSGIKAGAIISNFLPNNGRQF	1359

RESULT 3  
US-11-270-796-3  
; Sequence 3, Application US/11270796  
; GENERAL INFORMATION:  
; APPLICANT: Dong, Jianli  
; TITLE OF INVENTION: TREATMENT OF CANCER BY SIMULTANEOUS INHIBITION OF BRAF AND  
; TITLE OF INVENTION: RESTORATION OR MIMICRY PF p16 Ink4a ACTIVITY  
; FILE REFERENCE: 02420/1201581-US1  
; CURRENT APPLICATION NUMBER: US/11/270, 796  
; PRIOR FILING DATE: 2005-11-08  
; PRIOR APPLICATION NUMBER: 60/592,592  
; PRIOR FILING DATE: 2004-07-30  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 3  
; LENGTH: 1333  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-11-270-796-3

Query Match 78.3%; Score 5407; DB 6; Length 1333;  
Best Local Similarity 77.2%; Pred. No. 0;  
Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

Qy	4	AGLRGWLWALLRLAQSEPYTTHIQPGYCAFIDECKQNPGLSGSLMTLSNVSCLSNTPA	63
Db	3	AAWQGWLLWALLNSAQSLYTPTHKAGCTTYECCGNPGLSGSLMTLSNVSCLSNTPA	62
Qy	64	RKITGDHLILLOKICPRLTYGN-TOACCSAQOLVSLASLSITKALLTRCPACSDNFVN	122
Db	63	RHVTGDHALLQRVCPRLTYGNPDYVACCTKQLVSLDSSLITKALLTRCPACSENFVS	122
Qy	123	LHCHNTCSPNQSLFINVTRVAQLGAGQLPAVVAYEAFYQHSFAEQSYDSCSRVRVPAAT	182

Db	123	IHCHNTCSPNQSLFINVTRVQDRPQQLPAVVAYEAFYQHSFAEKAYESCSRVRIPAAAS	182
Qy	183	LAVGTWCGVYGSALCNAQRLWLFQGGTNGCLAPLDITTEHLLEPGQAVGSIQPLNEGVAR	242
Db	183	LAVGSMCGVYGSALCNAQRLWLFQGGTNGCLAPLDITTEHLLEPGQALADGMKPLDGIKTP	242
Qy	243	CNESQGGDDVATCSCODCAASCPAIAARPOALDSTFYLGQMPGSLVLIIILCSVFAVVTILL	302
Db	243	CNESQGGDDVATCSCODCAASCPVIPPALRPSFYNGRMFGWLALIIIFAVFVLLSVVL	302
Qy	303	VGFVAPADKSKWDPKGTSLSDKLSFSTHTLLQOFTQCGWGTWASWPLTTLILVSVIP	362
Db	303	VYLRVASNRNKNKTAGSQEAPNLPKRRRFSPTVLTGRFFESWGTVRASWPLTLVLSFIV	362
Qy	363	VVALAAGLVFTLTTDPVELWSAPNSQARSEKAFHDOHFGPPFRPTNOVILTAPNRSYRY	422
Db	363	VIALSVGLTFTLTTDPVELWSAPKQARKEAFHDEHFGPPFRPTNOIFVTAKNRSYKY	422
Qy	423	DSLLGPKNFSGILDLDDLLLELLELQERLRLHQLQVMSPEAQRNISLQDICVAPLNPNNTSL	482
Db	423	DSLLGPKNFSGILSLDLAQELLELQERLRLHQLQVMSHEAQRNISLQDICVAPLNPHNTSL	482
Qy	483	YDCCINSLLQYFQNNRTLLLTANQTLMGOTSOVDWKDHFPLYCANAPLTFKGTALALSC	542
Db	483	TDCCVNSLLQYFQNNHTLLLTANQTLNGQTSLVDMKDHFLYCANAPLTYKGTALALSC	542
Qy	543	MADYGAPVPELAIAGYKGDYSEALJMTSLNNYPAGDPRLAQAKWEERAFLEBMA	602
Db	543	IADYGAPVPELAIAGYKGDYSEALJITTSINNYPADDPMAHAKWEERAFLEBMA	602
Qy	603	FORMMAGFQVFTTAERSLEDEINRTTABDLPFATSYIVFIYISIALGYSWSRVMV	662
Db	603	FORSTADKQIAFSAERSLEDEINRTTIQDLPVFAISYIVFIYISIALGYSRNRVAV	662
Qy	663	DSKATLGLGGVAVVLGAVMAAGFPFSGYLSIRSLVILQVVPVLVSVGADNIFIFVLEYQ	722
Db	663	DSKATLGLGGVAVVLGAVMAAGFYSGYLSVGPSSVLIQVVPVLVAVGADNIFIFVLEYQ	722
Qy	723	RLPRPGBPREVHIGRALGRVAPSMLCSLSAI CFFLGALTMPMPAVRTRFALTSLGAVIL	782
Db	723	RLPRMPGEOREAHIGRTLGSVAPSMLLCSLSAI CFFLGALTSMPAVRTRFALTSLGAIIF	782
Qy	783	DELLQMSAFVALLSDSKRQASRLDVCCVKPQLPPGQEGGLLGFQKAYAPFLH	842
Db	783	DELLQMTAFVALLSDSKRQASRPDVCCPSSRNLPPEKQEGGLLCCFRKIYTFLLH	842
Qy	843	WITRGVLLFLALFGVLSYMSCHISVGLDQELALPKDSYLLDYFLNRYFEVGPVYF	902
Db	843	REFIRPVLLFLVLFGANLYLNCNITSVGLDQDLALPKDSYLLDYFLNRYLEVGPVYF	902
Qy	903	VTTLGYNFSSEAGMNAICSSAGCANNFSTQKIQYATEPPEQSILAIPASSWDDFDWLT	962
Db	903	DTTSGYNFSTEAGMNAICSSAGCESFSTQKIQYASEFPNQSYVAIAASSWDDFDWLT	962
Qy	963	P-SSCCRLYISGPNKDKFCPSTVNSLCLNCKMSITMGSVRPSVEQFHKYLWFLNDRPN	1021
Db	963	PSSSCCRITYTRGPHKDEFCPSTDTSFNCLNCKMNTLGPVTRTTTQFHKYLWFLNDRPN	1022
Qy	1022	IKCPKGLAAYSTSVNLTSDGOVLASRFMAYHKLKNSQDYTEALRAARELANITADLR	1081
Db	1023	IRCPKGLAAYVTSVNLSSDQIIASQFMAYHKPLRNSQDPTEALRASRLAANITAEALR	1082
Qy	1082	KVPGTDPAPFVPPYITTNVFEQYLTILPEGLFMLSCLVPTFAVSCULLGLDLSGLN	1141
Db	1083	KVPGTDPAPFVPPYITSNVFFYQYLTILPEGIFTLALCFVPTFVVCYLLGLDLSGLN	1142
Qy	1142	LLSIWMLVDTVGFMAWDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTWLERAA	1201
Db	1143	LLSIIMILVDIIGLMAVWGISYNAVSLINLVTAGMSVEFVSHITRSPAISTKPTWLERAA	1202
Qy	1202	KEATISMGSAVAGVAMTNLPGLVLGLAKAQILQIPFRNLNLTLLGLHLGVFLPVY	1261



Db 1203 KDATIFMGSAVAGVAMTNPFGIILILGPAQALIQIFFRILNLLITLGLLHGLVFLPVV 1262  
Qy 1262 LSYGPDVNPALALEOKBAEAAVAVMVASCPNHPSTADNIVVNHSPBGS-1KGAGA 1320  
Db 1263 LSYGPDVNPALALEOKBAEAAVAVMVASCPNHPSTADNIVVNHSPBGS-1KGAGA 1321  
Qy 1321 ISNFLPNNGROF 1332  
Db 1322 ASSSLPKSDQKF 1333

RESULT 4  
PCT-US06-05584-822  
; Sequence 822, Application PC/TUS0605584  
; GENERAL INFORMATION:  
; APPLICANT: DANA-FARBER CANCER INSTITUTE, INC.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION,  
; TITLE OF INVENTION: ASSESSMENT, PREVENTION, AND THERAPY OF CANCER  
; FILE REFERENCE: DFS-064.25(25992-6425)  
; CURRENT APPLICATION NUMBER: PCT/US06/05584  
; CURRENT FILING DATE: 2006-03-02  
; PRIOR APPLICATION NUMBER: 60/690,064  
; PRIOR FILING DATE: 2005-06-13  
; PRIOR APPLICATION NUMBER: 60/654,227  
; PRIOR FILING DATE: 2005-02-17  
; NUMBER OF SEQ ID NOS: 848  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 822  
; LENGTH: 1278  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US06-05584-822

Query Match 34.8%; Score 2402.5; DB 1; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 1.5e-185; Indels 103; Gaps 25;  
Matches 522; Conservative 237; Mismatches 450;

Qy 7 RGLWALLRLRLAQSEPTTTHQPGYCAFYDECG-----KNPELGSGLMTLSNVCSLS 59  
Db 4 RGLALGLLLLLCPAQVFSG-----SCVWYEGCIAYGDKRYNCEYSG-----46

Qy 60 NTPARKITGDHLILLOKICPLRYTGNTQACSAKQVLSLEASLSTIKALLTRCPACSDN 119  
Db 47 --PPKPLPKDGYDLVQELCPGFFG-NVSLCCDVRQLTKDNLQPLQLFLSRCPSCFYN 103

Qy 120 FVNLHCHTCSNQSLEFINVTR---VAOLGAGQLPAVVAEAFYQHSFAEQSDYSCSRV 175  
Db 104 LMLFCELTCSRQSQFLNVATEDYVDPVNTQKTNVQELQYVYVQSQFANAMYNACRDV 163

Qy 176 RVPAAATLAVGTGCGYVGSALCNAQRWLNFGQDTGNGLAPLDIT-----PHLEPGQA 228  
Db 164 EAPSSNDKALGLLCGKDADA-CNATNWIETFMKONGQAPFTITPVFDFPVH-----215

Qy 229 VSGIQPLNEGVARCNESQDDVATCSQDCAASCFAIARPO-----ALDSTFYLG 279  
Db 216 ---GMBPMNATKCDSEVDEVTAPCSQDCSVCGPKQPPPPPPAPWTLGLDAMVYIM 272

Qy 280 QMPGSLVLIILCSVPAVVTILL-----VGRFVAPARDKSNVDPKKGTSLS 326  
Db 273 WITYMAFLVFFGAFVAVCYRKYRYFVSEVTPIDSNIAFSV-NASDKGE-----ASCC 324

Qy 327 DKLSFSTHLLGQFPQGWGTWASWPLTILVLPVVALAAGLFTTELTDVPELWSAP 386  
Db 325 DPVSAAPGELRLRFLRWGSCVRNPGCVIFFSLVFTACSSGLVFRVVTNPDVLSWAP 384

Qy 387 NSQARSEKAPHDHGFQFPFTNQVILLTAPNRSYRVDLSLLGPK-NFSGILDLDLLELL 445  
Db 385 SSQARLEKEYFDHGFQFPFTNQVILLTAPNRSYRVDLSLLGPK-NFSGILDLDLLELL 444

Qy 446 ELQERLRLHQLWSPQEAQRNLSQDICVAPLNPNTSLYDCCNLSLLQYFQNNRLLTLLTA 505  
Db 445 DLQIAIEN--ITASYNDETTLQDICLAPLSPYNT---NCTILSVNLVFNQNSHVLDDHK 499

Qy 506 NQTIMGQTSQVDKWDHFLYCANAPLTFKQCTALALSCMADYGAPVPPFLAIGCYKGDYS 565  
Db 500 GDDPFF---VYADYTHFLCYCRAPASLNDTSLLDHPCLTGTFGGFVFWLVLGGDDQNYN 556  
Qy 566 EAEALIMTEFLNNYPAGDPRLAQALWEEAFLEEMRAFORRMAGMFQVTTAERSLDEI 625  
Db 557 NATALVITFPVNNYNDTEKLQRAQWEKEFINFKYKN---PNLTISFAERSIEDEL 613  
Qy 626 NRTTAEDLPITATSYIVIFLYISLALGSYSSWRVMVDKATLGLGVAVVLGAVMAAMG 685  
Db 614 NRESDSVETVWISYAIMFLYISLALGHKICRRLVDSKVSGLGIAGILVLSVACSLG 673  
Qy 686 FFSYLGIRSLVILQVVPFLVLSVGADNIIIFVLEYQRLPRRPEPEVHIGRALGVAP 745  
Db 674 VFSYIGLPLTLVIEVIFPLVAVGVNDNIFILVQAYORDERLOETLQQQLGRVLGEVAP 733  
Qy 746 SMLLCSLSEACFPLGALTTPMPAVRTPALTSGLAVILDLLQMSAFVALLSLDSKROAS 805  
Db 734 SMFLSSFSSEVAFPLGALSVPVAVHTSLFAGLAVFDILLQITCTFVSLGLLDIKRQEN 793  
Qy 806 RLDVCCCKQPELPPQO-GEGLLLGPFQKAYAPFLHWHITRQVVLVLLFLALFGVLSYM 864  
Db 794 RLDIFCCVGAEDGTSVQASESCLFRFFKNSYSPLLLLKDMRPVIAIPVGVLSFSIAVL 853  
Qy 865 CHISVGLDDELALPKDSYLLDYFLFNRYFEVGAAPVYFVTTLGVNFSSEAGMNAICSSAG 924  
Db 854 NKVDIGLDQLSMDPDSYMDYFKSISOYLHAGPPVFLVEEGHDTISSKQNNVCGMG 913  
Qy 925 CNPFSFTQKIQYATEFPEQSYLAIPASSWVDDDFIDWLTLP--SSCRLYISGPNKDKFCPST 983  
Db 914 CNNDSLVQOIFENAAQLDNYTRIGFAPSSWIDDYDFWVKPQSSCCRV---DNITDOFCNAS 970  
Qy 984 VNSLNLKNCMSIT-MGSVRPSVEQHKYLPWFILNDRPNKICPKGGLAAAYTSVN--LTS 1040  
Db 971 VVDPAQVR-CRPLTPEGKQRPQGGDFMRFLFPLMSDNPKNPKCGKHGAAYSSAVNILLGH 1029  
Qy 1041 DGQVLASFPMAHYKPLKNSQDYTRALRAARELANITADLRKVPCTDPAFVFPVTTINV 1100  
Db 1030 GTRVGATYFMTYHTVLTQTSADFIDALKKARLIASNT-ETWINGS--AYRVFFYSFVY 1086  
Qy 1101 FVEQVLTILPGLFMLSCLVPTFAVSCLLGLDLRSLGLNLLSIVMLVDTVGFMAIWD 1160  
Db 1087 FVEQVLTIIDTIFNLGSLGAIFLVTWVLGCELWSAVIMCATIYMLVNMFGVMMWLWG 1146  
Qy 1161 ISYNASVLINLVASVMSVEFVSHITSFALSTPTWLERAKEATISNGSAVEAGVAMTN 1220  
Db 1147 ISLNAVSLVNLVMSGISEVFCSHITRAFTVSMKGSRYERAEALAHMGSSVFSGITLTK 1206  
Qy 1221 LPTILVLGLAKAQLIQIFFRILNLLITLGLLHGLVFLPVLSYVGPVDPNPA 1272  
Db 1207 FGGIVVLAFAKSQIFIFFRMYLAWLLGATHGLIFLPLVLLSYIGPSVKA 1258

RESULT 5  
US-11-191-274A-129  
; Sequence 129, Application US/11191274A  
; GENERAL INFORMATION:  
; APPLICANT: Applera Corporation  
; APPLICANT: Bruno DOMO  
; TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF  
; FILE REFERENCE: CL001536PROV  
; CURRENT APPLICATION NUMBER: US/11/191,274A  
; CURRENT FILING DATE: 2005-07-28  
; NUMBER OF SEQ ID NOS: 334  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 129  
; LENGTH: 1278  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-191-274A-129

Query Match 34.8%; Score 2402.5; DB 6; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 1.5e-185;





```
Db 445 DLQIAIEN--ITASYDNETVTLQDICIAPLSPYNT---NCTILSVLNYFQNSHVLDDHKK 499
Qy 506 NQTLMGQTSQVDWKDHFYLCANAPLTFKGTALALSCMADYAGAPVFPFLAIGYKGYK 565
Db 500 GDDFF--VYADYHTEFLYCVRAPASLNDTSLHDPCLGTGFGVFPVPLVGLGYDDQNYN 556
Qy 566 EAEALIMTFSLNYPAGDPLAQAALWEAEFLERAPFORRMAGMFQVTFATERSLEDEI 625
Db 557 NATALVITFPVNNYNDTEKLQRAQAEKEFINFVKYKN---PNLTISFTAESIEDL 613
Qy 626 NRTTAEPLPATSYIVIFLYISLALGYSYSSWSRVMVDKATLGLGVAVVLGVMAAMG 685
Db 614 NRESDDVFTVVISYALMFLYISLALGHKICRLLVDSKVSIGIAGILIVLSSVACSLG 673
Qy 686 FFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLEYQRLPRRPGEPREVIHGRALGRVAP 745
Db 674 VFSYIGLPLTLVIEVIFPLVAVGVNIFILVQAYQDERLQGETLDQQLGRVLGEVAP 733
Qy 746 SMLLCSISEALCFELGALTMPAVRTFALTSLGLAVILDLQMSAFVALLSLSKROEAS 805
Db 734 SMFLSSFSETVAFPLGALSVMVPAHTFSLFAGLAVFIDFLQITCFVSLGLDIKROEKN 793
Qy 806 RLDVCCVCKPQELPPGQ--GEGLLGFFQKAYAPFLHWTITRGVLLFLALFGVSLYSM 864
Db 794 RLDIFCCVGAEDGTSVQASESCLFRFFKNSYSPLLLKDWMRPIVIAIFVGLSFI AVL 853
Qy 865 CHISVGLDQELALPKDSYLLDYFLFNRYFEVGAPEVFTVTLGNFSSSEAGMNAICSSAG 924
Db 854 NKVDIGLQSLMPPDDSYMVDYFKSISQYLHAGPPVYFVLEBHDYTSKQNMVCGGMG 913
Qy 925 CNNSFTOKIQAYATEPPEQSYLAIPASSWVDDDFDLWLT--SSCCRLYISGNKDKFCPST 983
Db 914 CNNDLSVQQLFNAQAOLNRYTRIGFAPSSWIDDFVWVKPOSSCCRV---DNITDQFCNAS 970
Qy 984 VNSLNCCLKNCSIT--MGSVRPSVEQPHKYLFWFLNDRPNIKCPKGLAAAYSTSVN--LTS 1040
Db 971 VVDPACVR--CRLTPGKQRPQGGDFMFLPWLSDNPNPKGKGHAAYSASVAILLGH 1029
Qy 1041 DGOVLASRWAYHKPLKNSQDYTEALRAARELAANITADLRKVPDPAFEVFPYITNV 1100
Db 1030 GTRVGATYFMTYHTVLQTSADFIDALKKARLIASNT--ETMGINGS--AYRFPYSFV 1086
Qy 1101 FFEQYLITPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSVIWLVDVTGFMALWD 1160
Db 1087 FFEQYLITLIDDFNLGVSGLAIFLVTWVLLGCELWSAVIMCATLAWLVNMFVGMWLMG 1146
Qy 1161 ISYNAVSLNLSVAGMSVEFVSHITRSPAITKPTWLERAKEATISMSGSAVFAGVAMTN 1220
Db 1147 ISLNAVSLNLSVMSCGISVEFCSHITRAFTVSMKGSVRERASEALAHMSSSVFSGITLTK 1206
Qy 1221 LBCILVLGAKAQLQIPIFFRLNLLITLIGLHGLVFLPVILSYVGPVNPA 1272
Db 1207 FGGIVLAPAKSQIQQIPIFFRYMLAWMLVGLGATHGLIFLPLVLLSYIGPSYNKA 1258
```

## RESULT 7

```
US-11-385-692-2400
; Sequence 2400, Application US/11385692
; GENERAL INFORMATION:
; APPLICANT: JOSELOFF, Elizabeth et al.
; TITLE OF INVENTION: KIDNEY DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001591ORD
; CURRENT APPLICATION NUMBER: US/11/385,692
; CURRENT FILING DATE: 2006-03-22
; NUMBER OF SEQ ID NOS: 6044
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2400
; LENGTH: 1278
; TYPE: PRN
; ORGANISM: Homo sapiens
US-11-385-692-2400
```

```
Query Match 34.8%; Score 2402.5; DB 7; Length 1278;
Best Local Similarity 39.8%; Pred. No. 1.5e-185;
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

Qy 7 RGMILLWALLLRAOSEPVYTTIHQPCYCAFYDECG-----KNPELSGLTSLNVSCL 59
Db 4 RGLALGLLLLLCPAQVFSQ-----SCVWYEGCGLAYGDKRYNCEYSG----- 46
Qy 60 NTPARKITGDHLILLOKICPLRYTGPNTQACCSAKQLVSLBASLITKALLTRCPACSDN 119
Db 47 --PKPLPKDGYLDVQLCEPFFFG--NVSLCDDVQLQTLKDNLQLPLQFLSRCPSCFYN 103
Qy 120 FVNLHCHNTCSPNOSLFINVTR-----VAQLGAGQLPAVAYEAFYQHSFASQSDSCSRV 175
Db 104 LNLFLCELTCSPROSQFLNVTATEDYDVPVNTQTKTNVKELOQYVYGOSFANAMYNACRDV 163
Qy 176 RVPAATLAVGTMGVYGSALCNAQRWLNFGDGTGNGLAPLIDIT-----PHLEPQA 228
Db 164 EAPSSNDKALGLLCKGQADA--CNATNLEYMFKNDGQAPFTITPFSDFPVH----- 215
Qy 229 VSGIQPLNEGVARNCSQGGDDVATCSQDCAASCAPAIARPO-----ALDSTFYL 279
Db 216 ---GMPEMNNATKGCDSVDEVTAPCSQCDSIVCGPKPQPPPPAPWTLGLDAMYVIM 272
Qy 280 QMPGSLVLIILCSVEAVVTILL-----VGFRVAPARDKSKMVDPKKGTSL 326
Db 273 WITNAFLVLPFGFAVAVCYRKRYFVSEYTPIDSNIAFSV--NASDKGE-----ASCC 324
Qy 327 DKLSEFSTHTLGGPFGQGTWVAWSPLTILVLSVIPVVALAAGLVFTELTTDPVELWSAP 386
Db 325 DPVSAAPFEGCLRLRFTRWGSCFVNPVCGVFFSLVFTACSSGLVFRVVTNPVLDMSAP 384
Qy 387 NSQARSEKAFHDQHFPGFPFRTNQVILFAPNRSRYRSLLLGPK--NFSGLDLDLLELL 445
Db 385 SSQARLEKEYDQHFPGFPFRTEQLIIRAPLTKHIYQPYPSGADVPPGLDIIQLHQL 444
Qy 446 ELQERLRLQVMSPEAQRNISLODICVAPLNPDNTSLDYCCINSLLOYFQNNRTLLLT 505
Db 445 DLQIAIEN--ITASYDNETVTLQDICIAPLSPYNT---NCTILSVLNYFQNSHVLDDHKK 499
Qy 506 NQTLMGQTSQVDWKDHFYLCANAPLTFKGTALALSCMADYAGAPVFPFLAIGYKGYK 565
Db 500 GDDFF--VYADYHTEFLYCVRAPASLNDTSLHDPCLGTGFGVFPVPLVGLGYDDQNYN 556
Qy 566 EAEALIMTFSLNYPAGDPLAQAALWEAEFLERAPFORRMAGMFQVTFATERSLEDEI 625
Db 557 NATALVITFPVNNYNDTEKLQRAQAEKEFINFVKYKN---PNLTISFTAESIEDL 613
Qy 626 NRTTAEPLPATSYIVIFLYISLALGYSYSSWSRVMVDKATLGLGVAVVLGVMAAMG 685
Db 614 NRESDDVFTVVISYALMFLYISLALGHKICRLLVDSKVSIGIAGILIVLSSVACSLG 673
Qy 686 FFSYLGRSSLVILQVVPFLVLSVGADNIFIFVLEYQRLPRRPGEPREVIHGRALGRVAP 745
Db 674 VFSYIGLPLTLVIEVIFPLVAVGVNIFILVQAYQDERLQGETLDQQLGRVLGEVAP 733
Qy 746 SMLLCSISEALCFELGALTMPAVRTFALTSLGLAVILDLQMSAFVALLSLSKROEAS 805
Db 734 SMFLSSFSETVAFPLGALSVMVPAHTFSLFAGLAVFIDFLQITCFVSLGLDIKROEKN 793
Qy 806 RLDVCCVCKPQELPPGQ--GEGLLGFFQKAYAPFLHWTITRGVLLFLALFGVSLYSM 864
Db 794 RLDIFCCVGAEDGTSVQASESCLFRFFKNSYSPLLLKDWMRPIVIAIFVGLSFI AVL 853
Qy 865 CHISVGLDQELALPKDSYLLDYFLFNRYFEVGAPEVFTVTLGNFSSSEAGMNAICSSAG 924
Db 854 NKVDIGLQSLMPPDDSYMVDYFKSISQYLHAGPPVYFVLEBHDYTSKQNMVCGGMG 913
Qy 925 CNNSFTOKIQAYATEPPEQSYLAIPASSWVDDDFDLWLT--SSCCRLYISGNKDKFCPST 983
Db 914 CNNDLSVQQLFNAQAOLNRYTRIGFAPSSWIDDFVWVKPOSSCCRV---DNITDQFCNAS 970
Qy 984 VNSLNCCLKNCSIT--MGSVRPSVEQPHKYLFWFLNDRPNIKCPKGLAAAYSTSVN--LTS 1040
```

Db 971 VVDPACVR--CRPLTPEGKQRQGGDFMRFLPMFUSDNPNPKCGRGHAAVSSA VNIILGH 1029  
Qy 1041 DGQVLASRFMAYHKPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPVPTITNV 1100  
Db 1030 GTRVGATVMTYHTVLTQTSADFDALKKARLIASNT-ETMGINGS--AYRVFPYSVFYV 1086  
Qy 1101 FVEQYLILPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSIVMILVDTVGFMAWD 1160  
Db 1087 FVEQYLTIIDDTFNLGVSGLGAIPLVTMVLGCELWSAVIMCATIAMVLNMFVGMWLMG 1146  
Qy 1161 ISYNAVSLINLVSAGVMSVEFVSHITRSEALSTKPTWLERAKEATISWGSNAVFAGVAMTN 1220  
Db 1147 ISLNAVSLNVMSCGISVEFCSHITRAFTVSMKGSVERAEALAHMGSSVFGITLTK 1206  
Qy 1221 LPGILVLGLAKAQILQIIFFRNLNLLITLGLLHGLVFLPVLTSYVGPDPVNP 1272  
Db 1207 FGGIVVLAFAKSQIFQIFVFRMYLAMVLLGATHGLIFLPLVLLSYIGPSVUKA 1258

RESULT 8  
US-11-385-692-2401  
; Sequence 2401, Application US/11385692  
; GENERAL INFORMATION:  
; APPLICANT: JOSELOFF, Elizabeth et al.  
; TITLE OF INVENTION: KIDNEY DISEASE TARGETS AND USES THEREOF  
; FILE REFERENCE: CL001591ORD  
; CURRENT APPLICATION NUMBER: US/11/385,692  
; CURRENT FILING DATE: 2006-03-22  
; NUMBER OF SEQ ID NOS: 6044  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2401  
; LENGTH: 1278  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-385-692-2401

Query Match 34.8%; Score 2402.5; DB 7; Length 1278;  
Best Local Similarity 39.8%; Pred. No. 1.5e-185;  
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

Qy 7 RGLLWALLILRLAQSEPYTHIOPGCAFVDECG-----KPELSGSLMTLSNVCSLS 59  
Db 4 RGLALGLLLLLCPAQVFSQ-----SCVMYGEIGAYGDKRYNCEYSG----- 46

Qy 60 NTPARKITGDHLILLOKICPLRYTGNTQACCSAKOLVSLASLSITKALLTRCPACSDN 119  
Db 47 --PPKPLPKDGYDLVDELCPGFFG--NVLCCDVRLQTLKONLQPLQLFLGRCPSCFYN 103

Qy 120 FVNLHCHNTCSNQSLFINVTR----VAQLGAGQLPAVVAYBAFYQHSFAEQSYDSCSRV 175  
Db 104 LNLPLCELTCSPQSQFLNVTATEDYVPTNQTKNVXELQYVYQCSFANAMYNACRDV 163

Qy 176 RVPAAATLAVGTWCGVYGSALCNAQRLNFGDGTGNGLAPLDT-----PHLLPQGA 228  
Db 164 EAPSSNDKALGLLGHKADADA-CNATWIEYMFNKONGQAPFTITPVFSDPPVH----- 215

Qy 229 VSGIOPLNEGVARCNESQDDVATCSQDCCAASCPAIPRQ-----ALDSTFYLG 279  
Db 216 ---GMEFPMNATKGDSEVDEVTAPCQDCQCSIVCGFKPQPPPPPPAPWTILGLDAMYIM 272

Qy 280 QMPGSLVLIIILCSVPVATVILL-----VGRVPAPARDKSKWDPKKGTSLS 326  
Db 273 WITYMAFLIVFFGAFVAVWCYKRYFVSEYTPIDSNIAFSV-NASDKGE-----ASCC 324

Qy 327 DKLSFSTHTLLGQFGGWTWASWPLTILVSVIPVVALAAGLVFTLTDPVLSWAP 386  
Db 325 DPVSAFAEGCLARLFRWGSFCVRNPGCVIFSLVFTACSSGLVFRVTTNPFVLDWSAP 384

Qy 387 NSQARSEKAFHQHFGPFRTHQVILTAPNRSYRDSLLLPK-NFSGILDLLLELL 445  
Db 385 SSQARLEKEYFOHFGPFRTHQVILTAPNRSYRDSLLLPK-NFSGILDLLLELL 444

Qy 446 ELQERLRHLQVMSPEAQRNISLQDICVAPLNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505  
Db 445 DLQIAIEN--ITASDYNETVTQLQDICLAPLSPYNT---NCTILSVLNYFNQSHSVLHKK 499

Qy 506 NQILMGQTSQVDKDHFLHCANAPLTFKGTALALSCMADYGAVPFPFLAIGGYKGDYS 565  
Db 500 GDDFF--VYADYHTFLCYVRAPASLNDTSLLDHPCLTGFGGPFVFWLVLGGYDDQYN 556

Qy 566 EABALIMTFLSNYPAGDPRIQAOKLWEAEAFLEMAFORRMAGMFOVTTAERSLDEI 625  
Db 557 NATALVITFPVNNYNDTEKLQRAQWEKEFINFNKYNKN---PNLTISTAERSIDEEL 613

Qy 626 NRTTABDLPIFATSYIVIFLYISLALGSYSSWSRVMVDKATLGLGGAIVVLGAAMAAG 685  
Db 614 NRESDDVFTVVISYAIMFLYISLALGHKISCRLLVDSKVSIGIAGILVLSSVACSLG 673

Qy 686 FFSYLGIIRSLVLTQVVPFLVLSVGDADNIFIVLEVQRLPRRPGEPREVIHIGALRVAP 745  
Db 674 VFSYIGLPLTLVIEVIFPLVAVGVDNIFILVQAYQORDERLQGETLDQQLGRVLGEVAP 733

Qy 746 SMLLCSLSEAIACFFLGCALTMPMPAVRTFALTSLGLAVILDFLQMSAFVALLSLDSKREAS 805  
Db 734 SMFLSFSSTVAFFLGALSVMFAVHTPSLFAGLAVFIDFLQITCFVSLGLDILKQKQEN 793

Qy 806 RLDVCCVKRQBELPPFGQ--GEGLLGFFQKAYAPFLHMTITRGVILLFLALFGLSVLSYM 864  
Db 794 RLDIFCCVGAEDGTSVQASESCLFRFFKNSYSPLLKDWMPRIVITAFVGLVSFSAVL 853

Qy 865 CHISVGLDQELALPKDSYLLDYFLFNRYFEVGAOPYVFTTGLYFNSSBAGMAICSSAG 924  
Db 854 NKVDIGLDQSLMPDDSYMVDYFKSISQYLHAGPPVYFVLEEGHDYTSKQGNMVCGMG 913

Qy 925 CNFSPKIOVATEPEEQSYLAIPASSWVDDFDLWLT-SSCCRLYISGPNKDKCPST 983  
Db 914 CNNDLSLVQOIFNAQLDNTIRIGFAPSSWIDYFDWMVKPQSSCCRV---DNITDQFCNAS 970

Qy 984 VNSLNCNLCMSIT-MGSVRPSVEQFHKYLPFLNDPRNPKCPKGLAAAYSTSVN--LTS 1040  
Db 971 VVDPACVR--CRPLTPEGKQRQGGDFMRFLPMFUSDNPNPKCGRGHAAVSSA VNIILGH 1029

Qy 1041 DGQVLASRFMAYHKPLKNSQDYTEALRAARELAANITADLRKVPGTDPAFEPVPTITNV 1100  
Db 1030 GTRVGATVMTYHTVLTQTSADFDALKKARLIASNT-ETMGINGS--AYRVFPYSVFYV 1086

Qy 1101 FVEQYLILPEGLFMLSCLVPTFAVSCLLGLDLRSGLNLLSIVMILVDTVGFMAWD 1160  
Db 1087 FVEQYLTIIDDTFNLGVSGLGAIPLVTMVLGCELWSAVIMCATIAMVLNMFVGMWLMG 1146

Qy 1161 ISYNAVSLINLVSAGVMSVEFVSHITRSEALSTKPTWLERAKEATISWGSNAVFAGVAMTN 1220  
Db 1147 ISLNAVSLNVMSCGISVEFCSHITRAFTVSMKGSVERAEALAHMGSSVFGITLTK 1206

Qy 1221 LPGILVLGLAKAQILQIIFFRNLNLLITLGLLHGLVFLPVLTSYVGPDPVNP 1272  
Db 1207 FGGIVVLAFAKSQIFQIFVFRMYLAMVLLGATHGLIFLPLVLLSYIGPSVUKA 1258

## RESULT 9

US-11-301-094-2  
; Sequence 2, Application US/11301094  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Diane J  
; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS  
; FILE REFERENCE: JB06242US01  
; CURRENT APPLICATION NUMBER: US/11/301,094  
; CURRENT FILING DATE: 2005-12-12  
; PRIOR APPLICATION NUMBER: 60/636,390  
; PRIOR FILING DATE: 2004-12-15  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 2  
; LENGTH: 1383

```
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-301-094-2

Query Match      19.4%; Score 1341.5; DB 6; Length 1383;
Best Local Similarity 27.5%; Pred. No. 2e-99;
Matches 369; Conservative 257; Mismatches 561; Indels 155; Gaps 40;

Qy 76 KICPRLVYGNWQCCSAKQVLSLEASITKALLTRCPACSDNFVNLHCHNTCSNQSL 135
Db 63 EFCPHLLTGDN-KLCCPTSAEGITKQIAQAHLIGRCPSCFDNPAKLWCEFTCSNQD 121
Qy 136 FINTRVAQL--GAGQLPAVAYEAF-----YQHS--FAEQSYDSCSRVRPAAATLAVG 186
Db 122 FVSIEMKPIEKKGFTPEYQPAEAYVNTVEIRLSTDAEGMFSCKDVTGGOPALRV- 180
Qy 187 TMCGVYGSALCNAQRWLNFOGDTGNGL-APLIDITFHLLPEQAVGSGIQP-LNBEVGARN 244
Db 181 -MC---TSTPCTLTNWLFEICTQLNDLNIPIHTKFLLYDPIKTPPSDRSTVMNVFTGCD 236
Qy 245 ESQGDVATCQCQCAASCPAIAQPALDSTFYLGOM-----PGSLVLIILICSVFAVVT 299
Db 237 KSARVGPACSTSEC--NKEEYANLIDLDGKTSQTCNVHGIACLNIFVMLAFIGSLAV 294
Qy 300 ILLYGF-----RVAPARDKSMVDPKKTSLSDKLSFSTHTLLGQFF 341
Db 295 LLCVGFVFTSDEEDVTNURQTSGEESPRNRK-----RTGANI-----HNFMENNA 342
Qy 342 QCWGTWASWPLTILVLSPVVALAAGLVFTELTDPVELWSAPNSQARSEKAPHQHF 401
Db 343 RDIGMAGNPKSHFFICGAVLIFCLPGMIYHKESTNVVDMWSSPRSRARQEEVFNANF 402
Qy 402 GPFFRTNVILTAPNRSRYRVDLLGPNKPSGILDLLELLELLELRLHQLWQSPSEA 461
Db 403 GRPQRYQIIMLL--SHRDFQSGKLYGP-----VFHKDIFBELFDILNAIKNIQTQSDG 455
Qy 462 QRNISLQICYAPLNPDTNSLYDCCINSILQYFQNNRTLLLTANOT-----L 509
Db 456 -RTITLDDVCVRPMFG-----YDCLIMPTNYFQGNKEHDMKSKETVSEDDDAFYF 510
Qy 510 MGQTSQVDKQHLFCANAPLTFKDGITALALSCMADYGAPVFPFLAIGYKGKQYSEAE 569
Db 511 SSEATTDEWNNHMAACIQPMHSQK--TKSGLSCMGTYGSPAPNM-VFGKNSTNHQAANS 567
Qy 570 LMTFSLNYPAGDPRLLAQAKWEAFLEEMRAPFORMMAGMFQVTFPAAERSLEIDNRTT 629
Db 568 IMMTILVTQ--RTEPEIQKAELEWEKFLKFCYREKSPKVI-FSFMABERSITDEIEN 624
Qy 630 AEDLPFATSYIVIFLYISLALGSY-----SSWSRVMDVSKATLGLGVAVVLGAVMAAM 684
Db 625 KDEIVTVVIALAFLGYVTFSLGRYFVCENQLWS--ILVHSRICLGMLSVINLLSSFCSW 683
Qy 685 GFPSYLGIRSSILVQVVPFLVLSVGADNIFIVLEY--QRLP--RRPGEPREVHIGRA 739
Db 684 GIFSMFGIHPKVNALVQVFPVTLGVCRTEFWVKYAAQQRVSMPSDQCPFI-VGMV 742
Qy 740 LGRVAPMLLCSLSEACFFLGPALTPMPAIVRTALTSLGLAVILDFLLQMSAFVALLSDS 799
Db 743 MAGTMPAMPSSSLGCAFSFFIGGFTDLPAIRTFCLYAGLAVLIDVVLHCTIIFALFVMDT 802
Qy 800 KQOEASRLDVCCKVPQELPPQCGRGL-----LLG-----FFQKAYAPFLHWHI 844
Db 803 QRELNG-----KP-EFFFPVQIKDLLGAVLIGRQATDTFMQTFHFQVAPFLMHRM 853
Qy 845 TRGVWLLFLALFGVSLYSWCHI SVGLDQELALPKDSYLLDYFLFLNRYFVEGAPYFVT 904
Db 854 TRIIITGIIIFASPIITVILSSKISVGFQDSMAFTKSYISTHRYLDKDFDVGVPVFPVTV 913
Qy 905 TLGYNFSSEAGMNAICSSAGCNFSPFTQKIQVATEPPEQSYLAIPASSWVDIDFWLT-P 963
Db 914 DGLDWHREDVQNKFTFPFGCSDTSGFNIMNAVAGHTQTYLSGENYMWIDYLEWISK 973
Qy 964 SSSCRLYISGPNKDKFCPSTVNSL-----NCLKNM-----SITMGSV---RPSVEQPH 1009

974 SPCKVYVHDPN--TFCSTNRNKSALDDKACRTCMDFDYVANSYPKSSIMYHRPSIEVFY 1031
1010 KYLPWELNDPRNLIKCPKGGIAAVSTVNLTSDQVLASRFMAYHKPL--KNSODYTEALR 1067
1032 RHURHFLEDPNSECVFGRASPKDAISFTSRGRIQASQWMTFHKKLSISNSDDFIKAMD 1091
1068 AARELAANTADILURKVGTPDAFEVFPYTTITNVFYEQYLTILPEGLFMLSLCLVPTPAVS 1127
1092 TARMVSRLERSI-----DDTAHVFAYSKIPFFEQYSTIMPILITLQTFITVVGVGII 1145
1128 CLLGLDLRSLNLNLISIVMILVDITGFMALWDISNVAVSLINLVSAGVMSVFEVSHITR 1187
1146 CVTLGIDVKGACAVICQVSNYFHVAFMYIFINPVNALSATNLVMSGGILIEFSVNVLK 1205
1188 SPATSKYPTWLERAKEATISMGSAVAGVAMTNLPGTLVLGLAKAQLIQIIPFRELNLIT 1247
1206 GYACSLRQRAKORAEESTVGSIGPIILSGPVVTWAGSTWFLSGAHLQIITVYFKLFLITI 1265
1248 LLGLLHGLVLPVILSYVGPVNPALALEQKRAEEAVAAVMVASCPNHPHRSVSTADNIYV 1307
1266 VSSAVHALIITLPIILAFGSGRHSSETSTNDNDQHDACVLS--PTAESHISNVERGIL 1323
1308 N-----HSFEGSINKGAGAI 1321
1324 NRPSLLDASHILDPLLKAEGGI 1345

RESULT 10
US-11-301-094-4
; Sequence 4, Application US/11301094
; GENERAL INFORMATION:
; APPLICANT: Levitan, Diane J
; APPLICANT: Smith, Marsha
; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS
; FILE REFERENCE: J06242U501
; CURRENT APPLICATION NUMBER: US/11/301,094
; CURRENT FILING DATE: 2005-12-12
; PRIOR APPLICATION NUMBER: 60/636,390
; PRIOR FILING DATE: 2004-12-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 4
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-301-094-4

Query Match      15.1%; Score 1046; DB 6; Length 1274;
Best Local Similarity 25.7%; Pred. No. 1.6e-75;
Matches 345; Conservative 241; Mismatches 587; Indels 168; Gaps 46;

Qy 10 LLWALLRLAQSEBPTTIHQGYCAFYDECGKNPSELGSLMTLSNVSCLSNTPARKITGD 69
Db 14 VLFLLLHLALCQ-----AKCM-TECDGEDSNHPPCKTNKSTYLPITVTRSLNPT 64
Qy 70 HLILLOKICPLXYTGN--TOACCSAKQLVLSLEASITKALLTRCPACSDNFVNLHCHNT 128
Db 65 YNARPEKYCSVLQOEDEKAQVCCTELQKGMTDRISNAATILGSCPCFDNFAXLWCQFT 124
Qy 129 CSPNQSLFINTRYVAQLGAGQLPAVAYEAFYQHSFAEQSYDSCSRVRPAAATLAVGTM 188
Db 125 CSPDQSKEMKWET-----TGPKNVVVKMEFKNRDFVEGLYESRHTWTFANGLALRLMSL 180
Qy 189 CGVYGSALCNAQRWLNFOGDTG-NCLA---PLDITFHLLPEQAVGSGIQPLNEGVARN 244
Db 181 CGKVS-----FENFYGFMTGKWLQASIPINTEFQFSRMKNANIPPTP-----CH 225
Qy 245 ESQGDVATCQCQDC---AASCPAIAQPALDSTFYLGQMPGSLVLIILCSVFAVVTIL 301
Db 226 KSAGKVPACGNAIDCTNAHQVLVDISKVEHLGTGVFHPFP-DFEWLLKICGLA-LTVL 283
Qy 302 LVGF-----RVAPARDKSMVDPKKTSLSDKLSFSTHTLLGQFFQG---WGTWAS 350
```

Db 284 LVFILKYSCHRSAPNGEDGCVYDLGKGN-----LEVOFEGLCARYANAVIK 330  
QY 351 WPLTILVSVIPVVALAAG-LVFTLTITDPVELMSAPNSQARSEKAF-HDQHFGRPPFRTN 408  
Db 331 HPLIFVSLGLIIVAAACCSGNFKFHSLSHVSVDQSAADGETRNEKFFHS--FGPNHRIE 388  
QY 409 QVILTAPNRSSVRYDSLLGPKNFSGILDLLELLELLEQLRHLQVMSPEAORNISIQ 468  
Db 389 QIFINLPPTT-----KSMFNPLFEEMFQLVGNIQNLT--ACYGNSVKLD 432  
QY 469 DICYPALPNDTSLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHPFLYCANA 528  
Db 433 DICYPKPIGKN-----HGCAIMSPNTYFQNKWTNFENAGPPTIDDEIFDDQHWHLKYCIEN 488  
QY 529 PLTFKDGKTALALSCMADYGAPVPPFLAIGG-----YKGD-YSEAEALIMTFSLNNYPAGD 583  
Db 489 PLTV--STYSEMSCFGEFGPDPLVFGGSNESIKGAEMYTARTIMITVIRG----- 541  
QY 584 PRLAQKWEAFLEEMRAFORRMAGMFQVTTFAERSLEDEINRTAED--LPFATSYI 641  
Db 542 PE-DOAIAWETAFLNMSRYEMKHANF---TFMTETSVAAEIHATAVETDKIVSVIACA 597  
QY 642 VIFLYISLALGSY--SSWSRVMDSKATLGLGGVAVLGVAVMAAGFSGYLGISSVLV 699  
Db 598 LIWVITMLGINHWPESSILSALVHHKLLISISAVMISVISWCSIGMFSLFGVHATDN 657  
QY 700 QVPEPLVLSVGDADNIFIFVLEYQRLPRRPGEP---REVH--IGRALGRVAPSMLLCSLS 753  
Db 658 VLVFVITCLGINRIFVIRTPQANGCHVCGLPNISYRENNHRSINMRRSIPIVLUNSLI 717  
QY 754 EAICFFL-GALTP-----MPAVRTFALTSGLAVIDLFLQMSAFVALLSLDSKRQ----- 802  
Db 718 CSTCLFLAGGLVYVSVSNPAVEVFARHAGLAILMDTAFYLIVMLPLFYQDARREWSKC 777  
QY 803 -----EASRLDVCCVKPQELPPQOGBGLLGFQKAYAFILHWHITRGVVLLFP 853  
Db 778 EIMPWYELSNESKINLCHAEVDGNLRSP-----VDWFKLAIAPELLLKICRIWIATFF 830  
QY 854 LALFQVLSYSMCHISVGLDQELALPKDSVLLDYFLFLNRYFEVGPAPVYFVTTLGYNFSE 913  
Db 831 FVSLIIACYCYLCLFEGFNQWAFSETSYLTGHFQNMNENLNGPPLMFVVEGDVKWHP 890  
QY 914 AGMAAICSSAGCNFPSTQKIQ---YATEFPEQSYLAIPASSWVDVDFIDWLTP--SSCRL 969  
Db 891 KWQNKFCFLAGDDNSMGNKIRSLAYAENY-KGNVHGDVNIWLDLSYLFQHPRCSCCKM 949  
QY 970 YISGPNKDKFC-PSTVNSLNCNKCMGTMGSVRPSVEQFHKYLFWFLNDRPNKCPKGG 1028  
Db 950 -----DGKQKCDPS--NATHC-SSCSSSVASLTTTEYEFYRNHLHFLETPPSIQCAHG 1001  
QY 1029 LAAYSTSVNLTSDGOVLASRFMAYHKP--LKNSDQYTEALRAARELANITADLRKVPGT 1086  
Db 1002 MALAKPALNLTNGKIQAYFSTFPKUNLSDSLQLYDAMWPAKYLADDIEREL-EIPGV 1060  
QY 1087 DPAFEVFYTTITNVFYEQYLITLPEGLFMLSCLVPTPAVSCLLIGLGLDLRSLNLSIV 1146  
Db 1061 ----KVYVYSTFFPYEQYLTSTVTVLVLVLFVAFVTTISLFLRVNLAGSLTVFVLL 1116  
QY 1147 MILVDTVGFMALWDISYNAVSLINLVSAVGMSVFVSHITRSFAISTKPTWLERAKRATI 1206  
Db 1117 SSVLHLEMYWLLGTTNVVSVINMAMSLGIAVEFFGQMLHGFVNSKKPKREERAFALV 1176  
QY 1207 SMGSVAVGAVMTNPLVGLV---LGLAKAQLIQIFPRNLNLTLLGLLHGLVFLPV 1262  
Db 1177 SNGSTTUSGI-----FPALMITAGCLSPADSRLVITYFCNLQVLGVCVAVGVVYMP 1232  
QY 1263 SYVGPDVNPALALBQKRAEEA 1283  
Db 1233 AIFGSDFFQNVSSSEESTDEA 1253

RESULT 11

US-10-953-349-21065  
; Sequence 21065, Application US/10953349  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nickolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDE  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 21065  
; LENGTH: 419  
; TYPE: PRT  
; ORGANISM: Glycine max  
US-10-953-349-21065  
  
Query Match 11.3%; Score 782.5; DB 6; Length 419;  
Best Local Similarity 38.4%; Pred. No. 7,1e-55;  
Matches 164; Conservative 88; Mismatches 138; Indels 37; Gaps 10;  
  
QY 863 SMC-HISVGLDQELALPKDSVLLDYFLFLNRYFEVGPAPVYFVTTLGYNFSSA-GMAIC 920  
Db 2 ALCTRIEAGLEQQIALPRDSYLGQYFSNISRYLVRGPPLYFYVK-DYNYLSKHTNQLC 60  
QY 921 SSAGCNFSTQKIQYATEFPEQSYLAIPASSWVDVDFIDWLTPS--SCRLYISGPNKDK 978  
Db 61 SISHCDSNLSLNEISRASLVPTSSYIAKPAASWLDLFLWISPEAFSCCRKF-----TNS 116  
QY 979 FCPSTVNSLNCNKCMSTMGSV-----RPSVEQFHKYLPMFLNDRPNI 1022  
Db 117 YCPDDQPPCCLPDBEGPCGLGVCKDCTTCFRHSDLVNDRPSTAQFREKLFWFLDALPSA 176  
QY 1023 KCPKGLAAYSTSVNLT--SDGOVLASRFMAYHKRPLKNSQDYTEALRAARELANITADL 1080  
Db 177 DCAKGHGAYTNSVDLNGYEGGVIQASEFRTYHTPLNRQGVYVNAIRAARDFSAISSSL 236  
QY 1081 RKVPGTDPAPFVPVYTTITNVFYEQYLITLPEGLFMLSCLVPTPAVSCLLIGLGLDLRSL 1140  
Db 237 K-----MDIFPYSPYIIFFEQYLDLWKLALNITVALGAFV-CLITSSVMSSAI 287  
QY 1141 NLSIVMILVDTVGFMALWDISYNAVSLINLVSAVGMSVFVSHITRSFAISTKPTWLER 1200  
Db 288 LLLVLMILDLMGVAILGILQNAVSVNLMISGIAVEFCVHIVHAFMWSLGD-RQR 346  
QY 1201 AKEATISMGSAVFAVAMTNLPGILVLGLAKAQLIQIFPRNLNLTLLGLLHGLVFLPV 1260  
Db 347 AKTALCTMGASVFGSITLTKLVGLVLCFSTSEIFVYVYFQMYLAVIIGFLHGLVFLPV 406  
QY 1361 ILSYVGP 1267  
Db 407 VLSLFGP 413  
  
RESULT 12  
US-10-461-673-16739  
; Sequence 16739, Application US/10461673  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Zhou, Ping  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Awang, Jian-Rui  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Zhang, Jie  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Xue, Aidong  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Wang, Zhi Wei  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Goodrich, Ryle W.  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Yamazaki, Victoria

Mon Apr 10 07:16:15 2006

APPLICANT: Utiwal, Manusha L.  
APPLICANT: Ma, Yunging  
APPLICANT: Chen, Rui-Hong  
APPLICANT: Ghosh, Malabika  
APPLICANT: Weng, Gezhi  
APPLICANT: Haley-Vicente, Dana  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 823  
CURRENT APPLICATION NUMBER: US/10/461,673  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: PCT/US02/29964  
PRIOR FILING DATE: 2002-09-19  
PRIOR APPLICATION NUMBER: US 10/245,014  
PRIOR FILING DATE: 2002-09-16  
PRIOR APPLICATION NUMBER: US 60/323,739  
PRIOR FILING DATE: 2001-09-19  
PRIOR APPLICATION NUMBER: PCT/US02/29636  
PRIOR FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: US 10/245,817  
PRIOR FILING DATE: 2002-09-16  
PRIOR APPLICATION NUMBER: US 60/323,349  
PRIOR FILING DATE: 2001-09-18  
PRIOR APPLICATION NUMBER: PCT/US02/29001  
PRIOR FILING DATE: 2002-09-13  
PRIOR APPLICATION NUMBER: US 10/243,552  
PRIOR FILING DATE: 2002-09-12  
PRIOR APPLICATION NUMBER: US 60/322,511  
PRIOR FILING DATE: 2001-09-13  
PRIOR APPLICATION NUMBER: PCT/US02/25485  
PRIOR FILING DATE: 2002-08-09  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 17116  
SOFTWARE: pf\_FL\_genes Version 6.0  
SEQ ID NO 16739  
LENGTH: 967  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-461-673-16739

Query Match 8.5%; Score 588.5; DB 6; Length 967;  
Best Local Similarity 21.3%; Pred. No. 1.4e-38;  
Matches 216; Conservative 192; Mismatches 359; Indels 249; Gaps 37;  
334 HT-----LLGQFFQGWGTWASWPLTILVSVVVALAAGLVF--TELTDPVELWSAP 386  
115 HTDLEGLSRFTQWLGWQGAHPWTFLLAPLMLTAALGTGLYLPKDEEDLEEHYTPV 174  
387 NSQARSEKAFHGHGFFRTNQVILITAPNRSS-----YRYSLLGLGKPNPGL 436  
175 GSPAKAERRFVQGH-----FTTNDYSRFSASRRSTEANFVSVLSYSDSLDPAFVYS 230  
437 DLDLLELLELQERLRLHQLWSPQAQRNLSLQDIC--YAPLPNDNTSLYDCCI--NSLQ 492  
231 KLDGAVQDDURV-----AREKSQIQYQVCARYAL-----CVPPNPILY 270  
493 YFQNRRLTLALTANQITMGQTSQVMDKDFLYCANAPLTPKQGTALALSCMADYGA-PVF 551  
271 AWQVNTLNL-----SSISFPAY-----NHGRHPLY 296  
552 PFLAIGY-----KQDYSEAEALIMTFSLANNYPAGDPRL-AQAKLWEEAFLEENRAF 603  
297 LTGFFGGYILGSLGMLGQLLRAKAMRLLYLK---TEDPEYDVQSKWMLTHLLDQFTNI 353  
604 QRMA-----GMFOVTFTAERSLEDEINRTTAEDLPFATSYIVFLYISL 649  
354 KNILALKIEVPGVGVLGGQGEKVHTLSRQLEFEATSVTVIPVFLAYILILPAVT 413  
650 ALGYSYSSWSRVMVDSKATLGLGGVAVVVGVAAMGFFSYLGRSSVLIVQVPELVLSV 709  
414 SCFRPD-----CIRNKMCAAFGVISAFVAVSGFGLLHIGV--PFVIVANSFPFLIGV 467  
710 GADNIFIVLEYQRLPRRPGPREVHIGRALGRVAPSMLLCSLEAICFFILGALTPMPAV 769

Db 468 GYDDMFIMISAWHK-TNLAGDIRE-RMSNVYSKAAVSITITITITNIALYTGIMSFPSV 525  
QY 770 RTFALTSLGLAVILDFFLQMSAFVALLSLDSKQEARSLDVCVCKVQPELP----- 819  
Db 526 QCFICYTGTTLFCYFYNYITFCGAFMALDGKEV-----VCLWLKADPKWPSFKKFC 580  
QY 820 -----PPQGEGL--LLGFFOKAYAPFLHWTIRGWVLLFLALPGVLSYSMCHSVGL 871  
Db 581 PFGSVPDDEHGTDIHPMSLFFRDYFGPFLTRSESKYFVVFVLYLIISSYIGCFHVG 640  
QY 872 DQELALPKDSYLLDYFLFLNRYF-EVGAPYVFTTILGYNPFSEAGNAICSSAGCNFSE 930  
Db 641 DLNRLASDDSYITPYFENVEENYFSDYGRVMVITVKVDYWDK-----DV 685  
QY 931 TQKIQVATEFPQSYLAIP--ASSWVDDFDIDMLTPSSCCRLYISGNPKDKFCFSTVNSLN 988  
Db 686 RQLENCCTKIFKKNVYVDKNLTFWLDAYVQ-----YLGKNSQD---PNEKNT-- 730  
QY 989 CLKNCMSITMGSVRPSVEQFHKYLPWFLNDRPNKCPKGLAAAYSTSVNLTSGQVLASR 1048  
Db 731 -----FMNIPDLSNFPN-----FOHDIINISSNEIISR 761  
QY 1049 FMAHKPLKNSQDYTEALRAARELANITADLAKVPCTDPAFVFPVITINVEQVLT 1108  
Db 762 GFITQTTDVSS-----AKKI---LLFQLRRI-AEDCQIPLMVNQAFIYFDQYAAI 809  
QY 1109 LPE-----GLFMLSCLVP-----TFVAVSLGLGLDLSGLNLLSIVMLV 1150  
Db 810 LEDTVRNVLVASAMFIVSULLIPYPLCSLWVTFAGSVIVGV----- 852  
QY 1151 DTVGFMALWDISYNAVSLINLVSAVGMVSFVSHITSFAISTKPTWLERAKEATISMG 1210  
Db 853 --TGFMAFKVNLDSISMINLVICIGFSDFSAHISYAFVSSQPSVQKSVREALYLLGY 910  
QY 1211 AVFAGVAMTNLPGLVLGLAKAQLIOIFFERLNLITLLGLLHGLVFLPVLSVYG 1266  
Db 911 PVLOS-AISTITIGVCVLAARAYIFRT-FFKIMFLVMIFGAHGLIPIFVPLTFPG 964

RESULT 13  
US-11-332-764-2  
; Sequence 2, Application US/11332764  
; GENERAL INFORMATION:  
; APPLICANT: Wisotzkey, Robert G.  
; TITLE OF INVENTION: PATCHED HOMOLOG 2 (PTCH-2) DISRUPTIONS, COMPOSITIONS AND METHOD  
; TITLE OF INVENTION: RELATING THERETO  
; FILE REFERENCE: R1673 CIP/75658.064000  
; CURRENT APPLICATION NUMBER: US/11/332,764  
; CURRENT FILING DATE: 2006-01-13  
; PRIOR APPLICATION NUMBER: US 60/413,543  
; PRIOR FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: US 10/669,143  
; PRIOR FILING DATE: 2003-09-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 1182  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-11-332-764-2

Query Match 8.4%; Score 583.5; DB 6; Length 1182;  
Best Local Similarity 24.2%; Pred. No. 4.7e-38;  
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;  
QY 330 SFSTHTLLGQ-----PFGQ-----WGTWASWPLTILVSVIPWVALAAGLVTELT 376  
Db 20 SSAPHTLAGSLQAPLWLRAYFQGLLFSLCGRICQKHCCKVLFLGLVAFGALGLRVAVIE 79  
QY 377 TDPELVWSAPNSQARSEKAFHGHG--PFFRTNQVILTPNRSSRYVDSLLLPKPSGI 435  
Db 80 TDLEQLWVEGVSQELHYTKELGEAAATTSQMLI-----QTAHOGGNNVLTPE-----A 131





302	TARDLQCLRAEALQSTFLMSPRQLYEHPRGDYQTHDIGNWBEQASQMLVQAWRRFPVQ	361
610	MFQVTFTRBSLE-----DEINRTTAEDLPFATSYIV--IFLYIISALGSYSSWS	658
362	LAQEALPANASQQIHAFSSTTDLDIRAFSE---VSTTRVVGYYLLMLAVCYACVMTLWRD	417
659	RVWVDSKATLGLGVAVVILGANVMAAGFFSVGLTRSSLVILQVVPFLVLSVGADNFI FV	718
418	--CAQSGQAGVLGAVLLVALAVASGLGCLALLGHTFNAANTQVLPFLALGIGVDDIFLLA	475
719	LEVORLPRRQCEPREVHIGRALGRVAPSMLLCSLEAICFFLGALTMPMPAVRTRFALTSG	778
476	HAFTKAP--PDTPLPERMGECLRSTGTISVALTSVNNVAFPMFMAALVIPALRAFSLOAA	533
779	AVILDFFLQWSAFVALLSLSOSKQRAERLNVCC-----VKPQE-----	817
534	VVGCNFAAVMLVFPAILLSLDLRRHRORQLDVLCCFFSPCSAQVQIOMQLPQELGDR	593
818	-----LPP-----PG-----QGBG--	826
594	AHLTATVQAFTHCBASQHVVTILPQAHLLSPASDPLGSELVSPGSGTRDILLSQEBGT	653
827	-----LLLGFFOKAYAPFLLHWHITRGVILLFLFALFGVLSYMSCHISVGLDQ	873
654	PQAAACRPLLCAHWTLAHFARYQYAPAPLQLQTRAKAVLLFFCALLGLSLYGATLVQDGLA	713
874	ELALPKDSYLLDYFLNRRYFEVCAPYVFTTILGYNES--SEAGMNAICSSAGCWNFSFTQ	932
714	TDVVPRGTGBHAFISLAQRYFSL--YEVALVTQGGFDYAHSQRAL-----FDLHQ	761
933	KIQVATEFFEQSYLAIPAS-----SWDDF-----IDWLTPSSCCRLYISGPNK	976
762	RFSSL-----KAVLPPATQAPRWLHYYSWLOGIQAAFDQDWSAGRITCHSVYNGSED	816
977	D----KFCPSTVNSLNCIKNCMSITM-----GSVRPSVEQFHXYLPWFNLDRPNKCPK	1026
817	GALAYKLLIQTGNAQEPDLFSQLTTRTKLVNKEGLIPP--ELFVGLTVVYSSDPL-----	869
1027	GGAAA-----YST--SVNLTSQGOVLASRFEMAYHKPLKNSODYTEALR	1067
870	GLAASQANFYPPPEWLHKYDITGTENLRIPAAQPLEPAQFPPLLHGLOKTADFEVRAIE	928
1068	AARELAANI-TADLRKVPGTDPAPFEVFPYTTINVFYEQYLTILPEGLFMLSCL--LVPTF	1124
929	GARAACTEAGQAGVHAYPSGSPF-----LFEQYLGCL--RRCFLLAVCILLVCTF	976
1125	AVSCLLILGLDLRSGLINLLSVMTLIVDTVGFMALWDISYNAVSLINILVSAVGSVEFVSH	1184
977	LVCALLLLSPWTAGLI-VLWAMTVBELFGIMGLGKLSAIPVWILVASIGIGVEFTVH	1035
1185	ITRSFAISTKPTWLERAKEATISMGSAVFAGV---AMTNLPGLIIVLGLAKAQITQIFPFR	1241
1036	VALGFLTSHGSRNLRAA-----SALEQTFAPTVDGAVSTLLGLULMAGSNFDIRFFV	1090
1242	INLLITLGLLHGLVFLPVILSYVGP	1267
1091	VLTVTLGLLHGLLPLVLSIIGP	1116

RESULT 15

US-60-772-265-1197

US-00-772-203-1137  
; Sequence 1197, Application US/60772265

; GENERAL INFORMATION:

APPLICANT: BOUKHAROV, ANDREY A.

APPLICANT: DU, ZIJIN

; APPLICANT: GUO, LIANG

APPLICANT: HRESKO, MICHELLE C

APPLICANT: KOVALIC, DAV

APPLICANT: ZHAOLONG, LI

; APPLICANT: LU, MAOLONG

APPLICANT: MCCARTER, JAMES P

APPLICANT: MILLER, NANCY  
APPLICANT: HAYDIN, MARK

APPLICANT: VAUDIN, MARK

```

; APPLICANT: WILLIAMS, DERYCK J
; APPLICANT: WU, WEI
; TITLE OF INVENTION: IDENTIFICATION AND USE OF TARGET GENES
; TITLE OF INVENTION: FOR CONTROL OF PLANT PARASITIC NEMATODES
; FILE REFERENCE: MNDI:002USP1
; CURRENT APPLICATION NUMBER: US/60/772,265
; CURRENT FILING DATE: 2006-02-10
; NUMBER OF SEQ ID NOS: 1919
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1197
; LENGTH: 891
; TYPE: PR1
; ORGANISM: Heterodera glycines
; FEATURE:
; OTHER INFORMATION: genomic DNA sequence=SeqID_396; coding sequence=SeqID_991
US-60-772-265-1197

```

Query Match	7.0%;	Score 482;	DB 8;	Length 891;
Best Local Similarity	22.2%;	Pred. No. 5.3e-30;		
Matches 203;	Conservative 160;	Mismatches 360;	Indels 190;	Gaps 35;
QY	540	LSCMADYGAPVFPFLAIGYK----	GKDYSEALIMTFSLNNYPAGD--	PRLAQAK--L 591
DB	10	MNCPED-----	LIVGIRDAQGVQIRHAEAFQTVFLVAS--	AGDYQELIRSKNHR 59
QY	592	WEBAFLIENRAFORRWAGMFQVTTFAERSLDEINRTTAEDL----	PIFATS-----	639
DB	60	MEKFGVSFTQAHAGATIAAWQRNFT--	KSIYDHTLNKPAQGIRVVHPLASTSIQDMLBQF	117
QY	640	-----YIVIEL--YISLALGSYSSWSRV-----	WVDSKATILGCGVAVVLGAVMAAMGFF	687
DB	118	SEQFFVIFIGVYLMII--	YAGWSQVHGWGWFVSKSCLLAIIGVITLIVLASVAGLGUS	175
QY	688	SYLGIRSSVILQVDPFLVLSVCADNIFIFVLEYQRLPRRPEPREVHIGRALGVAPSM	747	
DB	176	TAMNHFNAATQIVPFLTGLGIDDMFLLNHNDVLE--	AVRQKEVAVLLKETGMSV	232
QY	748	LLCSLSEATCFPIGALTMPAVRTFALTSGLAVIIDFLQMSAFVALLSLSKQREASRL	807	
DB	233	LITSINNILAFUTGCIPLPALRSFCGQVAILLTSNVLCIILLPFAFIALDURRERKAGHR	292	
QY	808	DVCCC-----VKPQELPPPGQGE-----	GLLLGFFQKAYAP	838
DB	293	DM5FCSSRNQLVKTDSTVTHTSSEIDLQMSAAVPTVSATKSLHKWYTLVGLFHYIYP	352	
QY	839	FLHLWITRQVVLILLFALFGVSLYSNCHITSVGLDQELALPKDSYLLDYFLFNRYFEVGA	898	
DB	353	LLRRPIAKTVLLVCAAMPLFCGFCGLYYSRIGLEADVLPEHTAPAAPLAKARERYFSP-Y	411	
QY	899	PVYFV---TTLGYNFSSEAGMNAICSSACGNPF-----	SFTQKI	934
DB	412	PMFIVPGKPMVDY--ANQCHKLEQLRRDIGRSNFVIKVDGEPBQWMMFMRTWLHLSQSSL	470	
QY	935	QYATE--PPEQSYLAIPASSMV-----	DDFIDWLTPSSCCRLYISGNKDKFCPSTVN	985
DB	471	DRAAKIGIFDDDLKXWADALARNFTLSDDFL--	LARKLLC-----	SAGQ 514
QY	986	SLNC--LKNCMSITMGSRVPSVEQFHXYL--	PMFLNDRPNIKCPKGG-----	LAAYS 1033
DB	515	RFNCTEMRGAKLIEDGRINP--	RGFNYLTAMFYQDNMMYYVVSQAAPFTPLPWHFSAAD	572
QY	1034	TSVNLISDQVQLASRPMATHKPKNSQDYTEALRAARELANITADLRKVPCTDPAPFVP	1093	
DB	573	ESVVPFAD--PLLYSQIPFYFMGLTDTQSTVQMIKBIACDRYSAD-----	GLPYY 622	
QY	1094	PVYITMVFYEOYLTILPEGLFMLSCLVPTFAVSCLLGLDLSRGLNLLSVMT--	LVD 1151	
DB	623	PSGIPPTFEQYVYKLT--	FYLVSALLIIGVAVLLVISIIIFNPWAAAMVAIVTMTVE	679
QY	1152	TVGFMALWDISYNVAVSLINLVSAVGMSVEFVSHITRSFAISTPKYPTWLERAKATTSMGSA	1211	
DB	680	LACFMGVFGVMNPTISAVTLITAVGIVGVEFTAHVLAFLTS-----	LASRDRERMVACLEH 734	



Qy	1212	VFAGV----	AMTNLPGLVLVLGLAKAQLIQI	FFRNLNLLITLLGLHGLVFLPVL	SVVGP-	1267
Db	735	MFIPVIHGGLSTLLGI	VMLAFSEBDFVVKYFFVVM	TALVIIGLGLALLPVL	SLIGPP	794
Qy	1268	-----	DVNPALALEQKRAEEA	VAAVMVASCNP	HP-----	SRVSTADNIYV 1307
Db	795	CEITPVNGSNLLPC	PASERYRPEDSPNC	FTTPSTHQQSD	EQKRRRGCKRMLAVDD---	851
Qy	1308	NHSFEGSIKGAGA	1320			
Db	852	SSSASSTEDGAGA	864			

Search completed: April 7, 2006, 19:33:58  
Job time : 83 secs

**THIS PAGE IS BLANK**

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE IS BLANK**